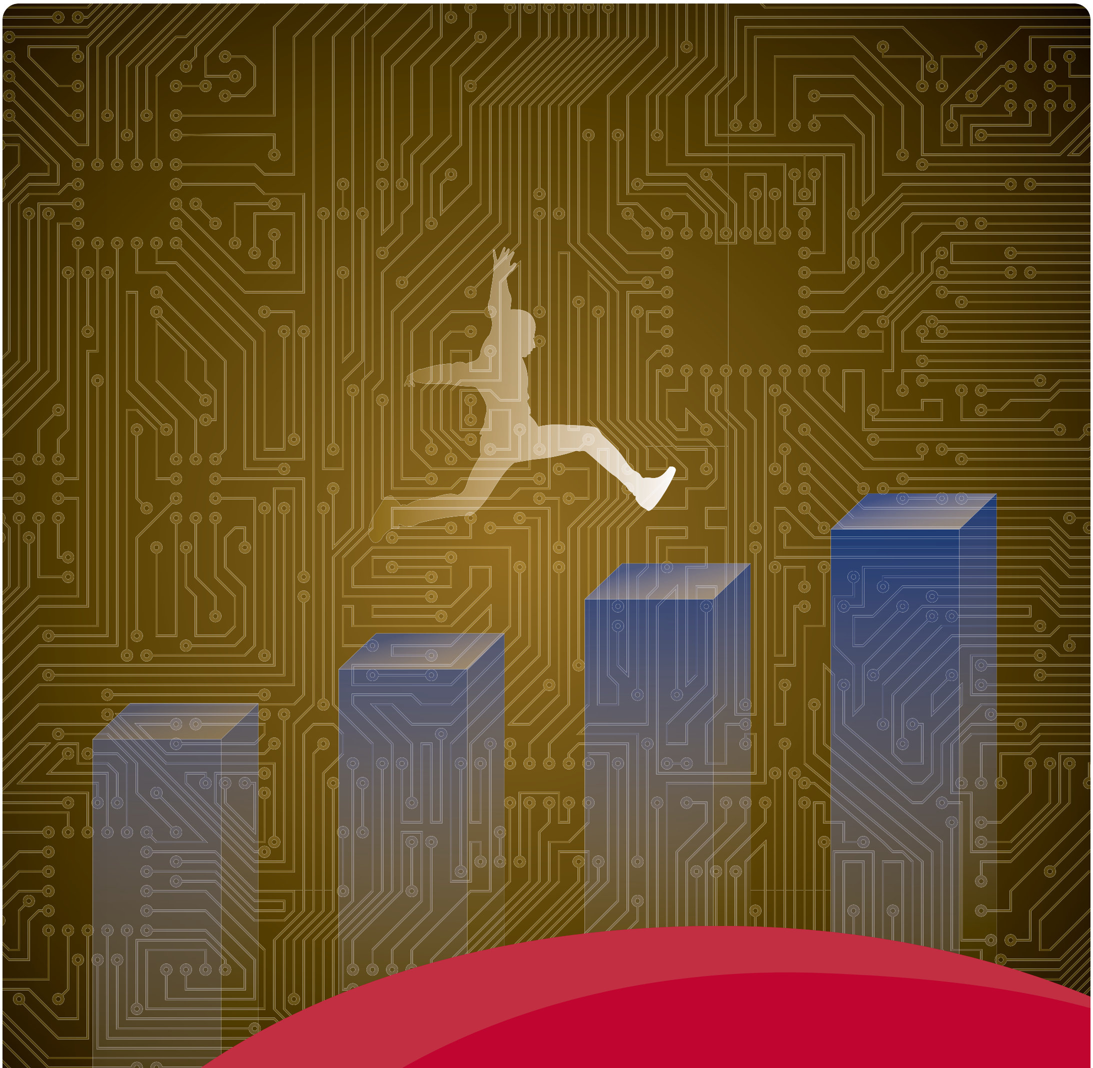


Leaps and Boundaries

The Expert Panel on Artificial Intelligence for Science and Engineering

CCA | CAC

Leaps and Boundaries

The Expert Panel on Artificial Intelligence for Science and Engineering

CCA | CAC

The Council of Canadian Academies
180 Elgin Street, Suite 1401, Ottawa, ON, Canada, K2P 2K3

The project that is the subject of this report was undertaken with the approval of the Board of Directors of the Council of Canadian Academies (CCA). Board members are drawn from the Royal Society of Canada (RSC), the Canadian Academy of Engineering (CAE), and the Canadian Academy of Health Sciences (CAHS), as well as from the general public. The members of the expert panel responsible for the report were selected by CCA for their special competencies and with regard for appropriate balance.

This report was prepared in response to a request from the National Research Council of Canada (NRC) along with the Canadian Institute for Advanced Research, Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council, and Social Sciences and Humanities Research Council. Any opinions, findings, or conclusions expressed in this publication are those of the authors, the Expert Panel on Artificial Intelligence for Science and Engineering, and do not necessarily represent the views of their organizations of affiliation or employment, or those of the sponsoring and co-sponsoring organizations.

Library and Archives Canada Cataloguing in Publication

ISBN: 978-1-990592-10-2 (book)
978-1-990592-11-9 (electronic book)

This report should be cited as:

Council of Canadian Academies, 2022. Leaps and Boundaries, Ottawa (ON). The Expert Panel on Artificial Intelligence for Science and Engineering, Council of Canadian Academies.

Disclaimer

The internet data and information referenced in this report were correct, to the best of the CCA's knowledge, at the time of publication. Due to the dynamic nature of the internet, resources that are free and publicly available may subsequently require a fee or restrict access, and the location of items may change as menus and webpages are reorganized.

This assessment was made possible with the support of the Government of Canada

The Expert Panel on Artificial Intelligence for Science and Engineering would like to acknowledge the Inuit, Métis, and First Nations Peoples who have been stewards of the lands now known as Canada.

The Council of Canadian Academies (CCA) acknowledges that our Ottawa offices are located in the unceded, unsurrendered ancestral home of the Anishinaabe Algonquin Nation, which has historically nurtured the land, water, and air of this territory and continues to do so today. Though our offices are in one place, our work to support evidence-informed decision-making has broad potential benefits and can hopefully contribute to collective action to address long-standing inequities and injustices impacting Indigenous Peoples. We are committed to drawing on a range of knowledges and experiences to inform policies that will build a stronger, more equitable, and more just society.

The Council of Canadian Academies

The Council of Canadian Academies (CCA) is a not-for-profit organization that supports independent, science-based, authoritative expert assessments to inform public policy development in Canada. Led by a Board of Directors and advised by a Scientific Advisory Committee, the CCA's work encompasses a broad definition of science, incorporating the natural, social, and health sciences as well as engineering and the humanities. CCA assessments are conducted by independent, multidisciplinary panels of experts from across Canada and abroad. Assessments strive to identify emerging issues, gaps in knowledge, Canadian strengths, and international trends and practices. Upon completion, assessments provide government decision-makers, researchers, and stakeholders with high-quality information required to develop informed and innovative public policy.

All CCA assessments undergo a formal peer review and are published and made available to the public free of charge. Assessments can be referred to the CCA by foundations, non-governmental organizations, the private sector, and any level of government.

www.cca-reports.ca

@cca_reports

The Academies

The CCA is supported by its three founding Academies:

The Royal Society of Canada (RSC)

Founded in 1882, the RSC comprises the Academies of Arts, Humanities and Sciences, as well as Canada's first national system of multidisciplinary recognition for the emerging generation of Canadian intellectual leadership: The College of New Scholars, Artists and Scientists. Its mission is to recognize scholarly, research, and artistic excellence, to advise governments and organizations, and to promote a culture of knowledge and innovation in Canada and with other national academies around the world.

The Canadian Academy of Engineering (CAE)

The CAE is the national institution through which Canada's most distinguished and experienced engineers provide strategic advice on matters of critical importance to Canada. The Academy is an independent, self-governing, and non-profit organization established in 1987. Fellows are nominated and elected by their peers in recognition of their distinguished achievements and career-long service to the engineering profession. Fellows of the Academy are committed to ensuring that Canada's engineering expertise is applied to the benefit of all Canadians.

The Canadian Academy of Health Sciences (CAHS)

The CAHS recognizes excellence in the health sciences by appointing Fellows based on their outstanding achievements in the academic health sciences in Canada and on their willingness to serve the Canadian public. The Academy provides timely, informed, and unbiased assessments of issues affecting the health of Canadians and recommends strategic, actionable solutions. Founded in 2004, CAHS appoints new Fellows on an annual basis. The organization is managed by a voluntary Board of Directors and a Board Executive.

Expert Panel on Artificial Intelligence for Science and Engineering

Under the guidance of its Scientific Advisory Committee, Board of Directors, and founding Academies, the CCA assembled the Expert Panel on Artificial Intelligence for Science and Engineering to undertake this project. Each expert was selected for their knowledge, experience, and demonstrated leadership in fields relevant to this project.

Teresa Scassa (Chair), Canada Research Chair in Information Law and Policy, Faculty of Law, University of Ottawa (Ottawa, ON)

Julien Billot, CEO, Scale AI (Montréal, QC)

Wendy Hui Kyong Chun, Canada 150 Research Chair in New Media and Professor of Communication, Simon Fraser University (Burnaby, BC)

B. Courtney Doagoo, Director, Management Consulting Practice, KPMG LLP; AI and Society Fellow at the Centre for Law, Technology and Society, University of Ottawa (Toronto, ON)

Abhishek Gupta, Founder and Principal Researcher, Montreal AI Ethics Institute (Montréal, QC)

Richard Isnor, Associate Vice President, Research and Graduate Studies, St. Francis Xavier University (Antigonish, NS)

Ross D. King, Professor, Chalmers University of Technology (Göteborg, Sweden); Professor, University of Cambridge (Cambridge, United Kingdom)

Sabina Leonelli, Professor of Philosophy and History of Science and Director of Egenis, University of Exeter (Exeter, United Kingdom); Fellow of the Wissenschaftskolleg zu Berlin (Berlin, Germany)

Raymond J. Spiteri, Professor, Department of Computer Science, University of Saskatchewan (Saskatoon, SK)

The CCA also recognizes the contribution of **Marc-Antoine Dilhac,** Professor (Philosophy), University of Montreal; AI CIFAR Chair; Director of Algora Lab; Co-Chair of Deliberation at the International Observatory of the Societal Impact of AI (OBVIA) (Montréal, QC).

Message from the President and CEO

Artificial Intelligence (AI) continues to capture the world's attention, in large part because it can undertake activities that could previously only be done by humans and has the potential to perform tasks humans could never do. To date, AI has been deployed alongside longstanding design and discovery practices to help researchers analyze or interpret data: for instance, AI has been used to predict the structure of proteins, identify chemical compounds for biomedical research, select preferred biomarkers for potential drug development, and track insect biodiversity.

Given how quickly AI continues to develop, it will very soon begin to play a more significant role in design and discovery for science and engineering. It's anticipated that AI will be used to develop novel scientific hypotheses and experiments, and create new engineering design processes, with less and less reliance on human programming.

Realizing the promise and potential benefits of AI will require addressing real and imminent challenges from possible biases, from the people who build it, to the institutions and governments whose policies are intended to regulate it. Ensuring the responsible use of AI may spur innovation and further scientific understanding, but there will be costs.

Recognizing the many complexities of AI's use across disciplines, the National Research Council of Canada and other supporting sponsors asked the CCA to examine the legal, regulatory, ethical, social and policy implications.

Leaps and Boundaries explores the opportunities, challenges, and implications of deploying AI technologies to enable scientific and engineering research design and discovery. The report identifies the actors whose decisions will determine how the challenges will be addressed and how various fields and sectors could potentially integrate AI into their practices.

Led by Chair Teresa Scassa, the Panel included members with expertise in law, ethics, humanities, applied science, industry, and policy. As with many of CCA's recent assessments undertaken during the COVID-19 pandemic, this panel met virtually online from start to finish. We thank them for the time, energy, and expertise that they put into this process. I also extend my thanks to the CCA's Board of Directors; Scientific Advisory Committee; and its founding Academies, the Royal Society of Canada, the Canadian Academy of Engineering, and the Canadian Academy of Health Sciences, for their guidance and oversight during the process.

Eric M. Meslin, PhD, FRSC, FCAHS
President and CEO, Council of Canadian Academies

Message from the Chair

Advances in artificial intelligence (AI) have the potential to transform the nature of scientific inquiry and lead to significant innovations in engineering. As a research tool that is increasingly used by more people and at more stages of both the design and the discovery processes, AI shifts the epistemic foundations of science and engineering. But the pitfalls of AI loom large in people's minds. If not used responsibly, AI could perpetuate human biases, and exacerbate inequities in the research system and society more broadly.

Addressing the social and ethical implications of AI in science and engineering, from the earliest stages of development through to deployment, will be critical to using it thoughtfully and responsibly. Establishing robust and transparent mechanisms to ensure the results generated by AI are accurate, reproducible, and explainable will also be key to realizing its benefits. Success will hinge on greater collaboration across disciplines.

In Canada, the AI ecosystem has focused on vertical growth but will need to grow horizontally beyond its existing strengths, crossing physical, disciplinary, and sectoral boundaries to maximize the opportunities in design and discovery in science and engineering.

Current legal and regulatory frameworks struggle with several novel challenges stemming from the use of an AI system for predictive or decision-making purposes, heightening risks in areas of trust and accountability. Governments and policy-makers will need to determine whether to adapt current legal frameworks to AI systems or develop new frameworks that address the uncertainties surrounding AI. Because many stakeholders are implicated in the development and deployment of AI for science and engineering, a central challenge for policy-makers is not only the need to develop new policies, but to coordinate them across disparate areas.

Integrating knowledge and skills across the social sciences, humanities, and health sciences will help to advance the understanding of AI for science and engineering, and will have implications outside the lab. Importantly, transdisciplinary collaboration can also help to address the longstanding equity, diversity, and inclusion issues associated with the use of AI in the Canadian research system.

It has been a pleasure to serve as Chair of this Panel. I would like to thank my fellow Panel members for their contributions and engagement through the process and the CCA staff for their steadfast support and guidance. Finally, I would like to thank the sponsors for submitting this question and making our work possible.

Teresa Scassa, SJD
Chair, Expert Panel on Artificial Intelligence for Science and Engineering

CCA Project Staff

Assessment Team:	**Jean Woo,** Project Director
	Alexei Halpin, Research Associate
	Matthew Ivanowich, Research Associate
	Weronika Zych, Senior Project Coordinator
	Karl Guebert, Research Consultant
	Tijs Creutzberg, Director of Assessments

With assistance from:

Designer	gordongroup
Editor	Jody Cooper Lisa Muirhead / Firefly Communications
Translator, En-Fr	François Abraham, traducteur agréé, Communications Léon inc.

Peer Review

This report was reviewed in draft form by reviewers selected by the CCA for their diverse perspectives and areas of expertise. The reviewers assessed the objectivity and quality of the report. Their confidential submissions were considered in full by the Panel, and many of their suggestions were incorporated into the report. They were not asked to endorse the conclusions, nor did they see the final draft of the report before its release. Responsibility for the final content of this report rests entirely with the authoring Panel and the CCA.

The CCA wishes to thank the following individuals for their review of this report:

Steven Berg, President and CEO, Aquanty Inc. (Waterloo, ON)

Ignacio Cofone, Assistant Professor (Law) and Canada Research Chair in Artificial Intelligence Law and Data Governance, McGill University (Montréal, QC)

Jason Edward Lewis, FRSC, Full Professor of Computation Arts, Design and Computation Arts; Concordia University Research Chair in Computational Median and the Indigenous Future Imaginary; Co-director, Indigenous Futures Research Centre (Montréal, QC)

Fred Gault, Professorial Fellow, United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (Maastricht, the Netherlands); Professor Extraordinary, Tshwane University of Technology (Pretoria, South Africa)

Maya Medeiros, Partner, Lawyer, and Patent Agent, Norton Rose Fulbright Canada LLP (Vancouver, BC)

Emanuele Ratti, Institute of Philosophy and Scientific Method, Johannes Kepler University (Linz, Austria); Department of Humanities and Arts, Technion, Israel Institute of Technology (Haifa, Israel)

Janna Rosales, Associate Professor (Teaching), Faculty of Engineering and Applied Science, Memorial University (St. John's, NL)

The peer review process was monitored on behalf of the CCA's Board of Directors and Scientific Advisory Committee by **Nicole A. Poirier, FCAE,** President, KoanTeknico Solutions Inc. The role of the peer review monitor is to ensure that the Panel gives full and fair consideration to the submissions of the peer reviewers. The Board of the CCA authorizes public release of an expert panel report only after the peer review monitor confirms that the CCA's report review requirements have been satisfied. The CCA thanks Dr. Poirier for her diligent contribution as peer review monitor.

Executive Summary

Artificial intelligence (AI) is enabling and enhancing scientific discovery in a variety of fields. It has the potential to drive future scientific investigation by allowing for automated hypothesis generation, experiment design, experimentation, interpretation, and analysis. From a technical standpoint, the development and deployment of AI systems for these purposes have been spurred by algorithmic advances, as well as access to vast and growing amounts of data. Given the scientific foundation of innovations that become widely used products and services, the implications of applying AI to the scientific discovery process are bound to be significant.

Recognizing the opportunities, challenges, and implications of applying AI to science and engineering, the National Research Council of Canada (NRC), with support from the Canadian Institute for Advanced Research (CIFAR), Canadian Institutes of Health Research (CIHR), Natural Sciences and Engineering Research Council (NSERC), and Social Sciences and Humanities Research Council (SSHRC), asked the Council of Canadian Academies (CCA) to convene an expert panel to answer the following question:

What are the legal/regulatory, ethical, social, and policy challenges associated with deploying artificial intelligence technologies to enable scientific/engineering research design and discovery in Canada?

In response to this request, the CCA assembled a multidisciplinary and multisectoral panel of experts in law, ethics, humanities, applied science, industry, and policy who met virtually five times over the course of 2021 and 2022 to evaluate the evidence and address its charge.

Report Findings

The capabilities of AI systems have grown and can be applied to an increasing variety of tasks in science and engineering

AI systems have a significant history in science and engineering, particularly as tools for analyzing data. However, AI is now also being applied to tasks further upstream in design and discovery due to technological advances in software and hardware development, unabated growth in the availability of data, and the emergence of industries that place these developments at the core of their operation. AI research tools not only improve the analysis of research results, but their capacity for pattern recognition and prediction can be used to identify new areas of research and development (R&D) for scientists and engineers. Complex systems and interactions may be studied more thoroughly by employing AI tools to render vast datasets more manageable. Indeed, AI applications are establishing new paradigms in some fields, such as drug discovery, and improving methods in others, such as software development. In the future, an AI system might take on a greater role in organizing research or proposing designs. This potential disruption extends to crucial tasks across the diverse workflows in science and engineering, from hypothesis generation to interpretation and analysis. These tasks could all be carried out autonomously by AI systems following advances in robotics and interfacing between machines and humans. Although this scenario is far from the present-day capabilities of AI systems — which still demonstrate numerous limitations — trends suggest that few fields will remain untouched by AI, warranting preparedness on the part of the R&D ecosystem.

The increased use of AI in science and engineering creates new epistemic, methodological, and ethical challenges for researchers

Although AI has the potential to improve reproducibility in science, it is currently hampered by a lack of transparency in sharing code and data, which ultimately undermines trust in the accuracy of results. There are ongoing efforts to promote transparency in models and benchmarking through shared datasets, as well as through disclosure requirements for publications and conferences. The creation of standards for data and metadata can facilitate increased transparency in this area, as can support for open data in research.

The possibility of AI systems producing inaccurate results also poses a risk for applications in science and engineering. AI systems may produce inaccurate or skewed results due to biases in training datasets and problems of generalizing from training data to new data. Several techniques help to identify and eliminate biases during the training phase, which can avoid unwanted outcomes and inaccurate results. While some forms of bias are unavoidable, developers can be trained to avoid bias in data collection and curation.

Even when the results generated by AI are accurate and unbiased, several popular types of AI systems operate as *black boxes*, such that it may be difficult or even impossible to explain how their results were generated. This can hinder scientific explanation and understanding and potentially undermine the credibility of machine-generated scientific findings. Although, in some cases, accuracy alone may be sufficient for scientific progress, the goal of science is ultimately explanation and understanding, and thus there are many reasons to prefer interpretable AI models over black boxes in science and engineering research.

Ethical considerations about the use of AI in science and engineering arise at all stages in the research process, including data collection and pre-processing; design and use of AI models trained on those data; dissemination and publication of results; and long-term storage, maintenance, and access to data, models, and results. AI will thus impact institutional policies around the responsible and ethical conduct of research. Furthermore, AI research complicates traditional notions of consent when human participants are involved, as it may rely on datasets containing information about individuals without their knowledge or consent. Data stewardship and management principles will need to be implemented to facilitate responsible and ethical data sharing and use. These include the well-known FAIR data principles, as well as other data management principles that complement FAIR, including TRUST, FACT, and — in the case of data involving Indigenous Peoples — CARE.

The R&D environment in Canada will face challenges adapting its practices for the assessment of research and researchers using AI

A growing number of countries, including Canada, have developed national strategies for advancing AI capacity. The Canadian strategy has focused historically on building a critical mass of AI investigators; more recently, it has begun targeting multiple sectors and considering societal impacts more explicitly. Broadening the scope of national AI strategies can be beneficial because new connections will need to be made in the R&D network in order to apply specialized AI knowledge to problems in science and engineering. Scientific research that uses AI may also dovetail with existing policy goals and initiatives, which might indirectly influence the trajectory along which AI develops.

Bolstering the role of AI for R&D in science and engineering will not be limited to laboratory settings. Impacts will be felt across the discovery and innovation landscape. The use of AI in the research process is blurring the boundaries among disciplines, rendering single-themed funding competitions, scientific reviews, and research programs less useful because domain expertise is unable to properly review the interdisciplinary nature of AI-driven science. As a result, the humanities and social sciences play an important role in scientific and engineering R&D that uses AI. Existing traditional disciplinary partitions in research funding may need to be rethought to ensure a fair and appropriate assessment of research using AI.

> "The use of AI in the research process is blurring the boundaries among disciplines, rendering single-themed funding competitions, scientific reviews, and research programs less useful because domain expertise is unable to properly review the inter-disciplinary nature of AI-driven science."

Given AI's promise for accuracy and consistency, the concept of researcher responsibility is subject to change. If AI systems become the standard for undertaking scientific tasks, expectations surrounding the responsibility for accuracy and completeness may eliminate work done by humans from consideration in research funding competitions. This could have the effect of removing applicants who lack access to AI from funding consideration. AI in the research process complicates issues such as reproducibility, explainability, and accuracy that would require an update to the Tri-Agency Framework for Responsible Conduct of Research.

The ability of AI to predict the impact of scientific research may be of use to governments and funding agencies making decisions about whether (and how) to provide funding opportunities to the scientific community. Additionally, it can support the peer review process used to evaluate prospective scientific applications. However, these potential uses will need to be carefully tested to ensure that they do not result in unintended consequences. For example, there are concerns that AI could exacerbate the marginalization of traditionally under-represented groups in the research system. Although the integration of AI into research funding and peer review systems could mitigate long-standing issues of equity, diversity, and inclusion, it cannot be presumed. Careful implementation, keeping in mind these limitations, will be necessary.

The increased deployment of AI systems for science and engineering risks perpetuating discrimination or biases both within the Canadian R&D environment and in broader society

For AI to be used responsibly in science and engineering, it should avoid perpetuating bias and discrimination against individuals or groups; yet AI tools have already been observed to amplify these issues in numerous real-world examples. The lack of gender and racial diversity in the field of AI research is well documented, and there are currently high levels of inequality in the existing distribution of resources, infrastructure, and skills in the context of the production, dissemination, and use of AI for scientific research. AI could expand this "digital divide," given the high cost of computational resources and increased competition, especially when public investment has primarily benefited the private sector rather than universities or the public sector. This possibility raises concerns about the monopolization of scientific knowledge. Moreover, because AI may be used for decision-making in scientific and engineering contexts — for example, for peer review and funding decisions — it has broader social implications, such as determining which problems are addressed by research.

The use of AI is likely to have wider social impacts on the science and engineering labour market, public trust in AI and science, the environment, and cybersecurity

Lack of trust in AI may act as a barrier to its adoption in science and engineering contexts. To overcome these barriers, those designing AI systems will need to build a trusted evidence base by transparently demonstrating successful, reproducible results. Lack of public trust in AI in other domains might negatively affect perceptions of the trustworthiness of AI for science and engineering, as could biased or discriminatory results or unethical practices around data collection and use. Trust in AI for science and engineering will also require addressing the security of AI systems and the risks to the owners, users, and those affected by such systems.

The increased use of AI could also impact the labour market in science and engineering. Some job displacement is inevitable; however, the primary effect of AI on scientific and engineering occupations is job transformation. Skills development and training will be needed to adapt the science and engineering workforce to the changes generated by the increased use of AI. This will require developing the technical knowledge and skills needed to innovate and deploy AI for research purposes, as well as providing future scientists and engineers with the ability to identify and address the social and ethical considerations associated with the development and use of AI.

More generally, tensions continue to grow surrounding the environmental impacts of AI. On the one hand, discoveries made using AI systems could help to address the climate crisis. On the other hand, the development and operation of AI systems require considerable amounts of energy and can produce significant greenhouse gas emissions. Furthermore, environmental impacts of AI also include those related to raw material extraction and manufacturing of components; transportation of materials; construction and installation of AI infrastructure; maintenance, repair, refurbishment, and upgrades to the system; and end-of-life stage, including transportation, waste processing, and disposal.

Technological development is outpacing the development of legal and regulatory frameworks that govern AI systems, leading to uncertainty with deployment and commercialization

In the legal domain, several hurdles are emerging around the use of AI in science and engineering, most acutely in the domain of intellectual property (IP). Traditional IP frameworks are difficult to apply to machines capable of innovation or originality. Patents and copyrights have become key assets in an increasingly digital and intangible economy; however, these instruments were originally conceived as incentives with humans in mind. This issue is particularly relevant to Canada, given recent initiatives to derive greater returns from Canadian innovation.

Ownership or control over data — as well as access to data — are emerging as key concerns for AI development. Policies facilitating access to data and open data and promoting responsible data governance and management are important for AI in science and engineering; however, regulatory gaps and the fragmentation resulting from the complex legal framework governing data in Canada may hinder access to certain types of data. Tensions also exist between transparency and secrecy, particularly from the standpoint of commercialization. Innovators may be cautious about sharing valuable data since they may be better protected through confidentiality rather than other formal IP protection mechanisms. Open data and data governance initiatives, combined with legislative reforms, may resolve some of these tensions, although developments are not necessarily coordinated.

Beyond IP, legal liability remains an additional area of ongoing debate. Current frameworks struggle with the attribution of responsibility for harms resulting from the use of an AI system for predictive or decision-making purposes, heightening risks in areas of trust and accountability. Governments and policy-makers will need to determine whether to adapt current liability frameworks to AI systems or new liability frameworks that address the uncertainties surrounding AI.

There is no single law for AI regulation or governance in Canada; the federal-provincial/territorial division of powers presents challenges to the creation of a single regulatory framework. Laws governing personal data at both the provincial/territorial and federal levels have begun to address the use of personal data in automated decision-making and may address anonymized data. Many activities in science and engineering are thus not currently regulated because they might not involve personal data. The resulting fragmentation presents a challenge given the collaborative nature of AI development, with implications for commercialization in the context of private–public partnerships and international partnerships.

Despite ongoing efforts to prepare Canada for economic competitiveness in a changing regulatory environment, there remain substantial challenges to develop and apply AI responsibly and ethically given uncertainty over operationalizing and accounting for the technology. Approaching AI systems as socio-technical systems may address such challenges by broadening the scope of consideration to include developments that may seem peripheral to the particular AI application at issue.

The use of AI systems in science and engineering are pushing disciplinary boundaries, collaboration, and coordination towards a transdisciplinary future

Although some legal/regulatory, ethical, social, and policy challenges that arise from using AI in science and engineering may be addressed by reconfiguring the technical systems themselves, this report suggests that increased transdisciplinary work will be necessary. A narrow approach to science and engineering — whereby science, technology, engineering, and mathematics disciplines remain separate from the social sciences, humanities, and health sciences — is increasingly inappropriate to advance understanding and explanation.

As a research tool that is increasingly used by more people and at more stages of the discovery process, AI shifts the epistemic foundation of science and engineering. Methods and procedures to ensure that the knowledge generated using AI is trustworthy — as well as accurate, explainable, and reproducible — will need revisions. Concerns about the transparency of AI systems will therefore need to be addressed. Furthermore, the integrity of Canada's research system will also depend on the responsible deployment of AI systems in ways that align with evolving expectations surrounding ethics, equity, diversity, and inclusion.

> "A narrow approach to science and engineering — whereby science, technology, engineering, and mathematics disciplines remain separate from the social sciences, humanities, and health sciences — is increasingly inappropriate to advance understanding and explanation."

When advancements made possible by AI systems being used in science and engineering are applied to the real world, they may improve some lives but also harm others. Legal and policy frameworks will need amending with AI in mind if safety and well-being are to be reconciled with innovation. IP issues concerning data and automated discovery also remain unclear. While contracts are currently being used to establish rights, responsibilities, liability, and so on among parties, regulatory reform will be necessary if there is to be some level of certainty for scientists and engineers using AI systems.

These challenges compel transdisciplinary thinking and collaboration. A national AI strategy is beginning to apply such an approach to transform Canada's research ecosystem by coordinating investment, policy reform, and training. A transdisciplinary approach would also support needed changes to research funding and education policies if Canada is to continue to promote ethics, equity, diversity, and inclusion in science and engineering.

Abbreviations

This list is not exhaustive of all the abbreviations used in the report but provides the reader with the most commonly used terms.

ADM	automated decision-making
AI	artificial intelligence
Amii	Alberta Machine Intelligence Institute
CIFAR	Canadian Institute for Advanced Research
CIHR	Canadian Institutes of Health Research
EDI	equity, diversity, and inclusion
GHG	greenhouse gases
IP	intellectual property
LESP	legal/regulatory, ethical, social, and policy
Mila	Quebec Artificial Intelligence Institute (formerly the Montreal Institute for Learning Algorithms)
NRC	National Research Council Canada
NSERC	Natural Sciences and Engineering Research Council
R&D	research and development
RDA	Research Data Alliance
RDM	research data management
SME	small- and medium-sized enterprise
SSHRC	Social Sciences and Humanities Research Council
TDM	text and data mining
XAI	explainable artificial intelligence

Glossary

Algorithm: "formula or set of rules (or procedure, processes, or instructions) for solving a problem or for performing a task. In Artificial Intelligence, the algorithm tells the machine how to find answers to a question or solutions to a problem. In machine learning, systems use many different types of algorithms. Common examples include decision trees, clustering algorithms, classification algorithms, or regression algorithms" (Guo *et al.*, 2019).

Artificial intelligence (AI): for the purposes of this report and to allow for a broader and more inclusive interpretation of AI technologies, AI is defined as a collection of statistical and software techniques, as well as the associated data and the social context in which they evolve. Furthermore, the term AI is used interchangeably to describe various implementations (methods) of machine-assisted design and discovery:

> **Deep learning:** "subfield of machine learning concerned with algorithms that are inspired by the human brain that works in a hierarchical way. Deep Learning models, which are mostly based on the (artificial) neural networks, have been applied to different fields, such as speech recognition, computer vision, and natural language processing" (Guo *et al.*, 2019).
>
> **Machine learning:** "normally refers to the branch of AI focused on developing systems that learn from data. Rather than being explicitly told how to solve a problem, [machine learning] algorithms can create solutions by learning from examples (referred to as "training" the [machine learning] algorithm)" (King & Roberts, 2018).
>
> **Reinforcement learning:** "type of dynamic programming that trains algorithms using a system of reward and punishment. The algorithm is exposed to a total [sic] random and new dataset and it automatically finds patterns and relationships inside of that dataset. The system is rewarded when it finds a desired relationship inside of that dataset but it is also punished when it finds an undesired relation. The algorithm learns from awards and punishments and updates itself continuously. This type of algorithm is always in production mode. It requires real-time data to be able to update and present actions. The agent learns without intervention from a human by maximizing its reward and minimizing its penalty" (Guo *et al.*, 2019).

Autonomous researcher (or AI scientist): “constellation of software and hardware modules dynamically interacting to accomplish tasks [...and] capable of autonomously carrying out research to make major scientific discoveries” (Kitano, 2021).

Bias:

Automation bias: occurs when people put trust into “automated support systems. This trust is the product of humans’ perception of these systems as having superior analytical capabilities than their human counterpart” (Lopez *et al.*, 2019). In contrast, **algorithmic aversion** occurs when researchers choose human predictions over algorithmically generated ones, even after seeing evidence that the human predictions are less accurate (Dietvorst *et al.*, 2015).

Biased datasets or models: when an AI system produces skewed or inaccurate results (including discriminatory results) due to a variety of factors, including biases in datasets resulting from pre-existing historical discrimination, sampling errors, or pre-processing of data; as well as from the subjective decisions made by researchers when developing the model, such as an inappropriate choice of model, model parameters, lack of human oversight, and lack of transparency, among others (Veale & Binns, 2017; WEF, 2018).

Discriminatory bias: in common usage and in many media reports, the term bias typically refers to AI models that produce discriminatory results (Hellström *et al.*, 2020); that is to say, when “unfair judgments are made because the individual making the judgment is influenced by a characteristic that is *actually* irrelevant to the matter at hand, typically a discriminatory preconception about members of a group” (Muller, 2020).

Inductive bias: one of several types of biases that are necessary for an AI to function. Inductive bias allows AI to generalize from its training data to new examples, because “[e]ffective learning from finite data requires assumptions about the data source ... given any finite amount of data generated from an unknown source, it is impossible to predict the next element, even approximately accurately, unless some prior information about the source is available” (Amit & Meir, 2019).

Black box: AI system whose inputs and outputs are known, but its inner workings are not understood. “Once trained [a black box AI system] can produce statistically reliable results, but the end-user will not necessarily be able to explain how these results have been generated or what particular features of a case have been important in reaching a final decision” (The Royal Society, 2019).

Data:

Big data: "large datasets that are produced in a *digital* form and can be analysed through *computational* tools" (Leonelli, 2020).

Data provenance: "examining the history and process of dataset construction, and considering how cultural norms and stereotypes were numerated and represented at the time of data creation" (West *et al.*, 2019). This is often used to manage and avoid potential biases in datasets.

Metadata: data that give information about other data.

Open data: "structured data that is machine-readable, freely shared, used and built on without restrictions" (GC, 2020a).

Ethics of AI: "sub-field of applied ethics, focusing on the ethical issues raised by the development, deployment and use of AI. Its central concern is to identify how AI can advance or raise concerns to the good life of individuals, whether in terms of quality of life, or human autonomy and freedom necessary for a democratic society" (AI HLEG, 2019).

Explainable AI/XAI: "efforts to make sure that artificial intelligence programs are transparent in their purposes and how they work. Explainable AI is a common goal and objective for engineers and others trying to move forward with artificial intelligence progress" (Guo *et al.*, 2019).

Indigenous Data Sovereignty: "ability for Indigenous peoples, communities and Nations to participate, steward and control data that is created with or about themselves. The term sovereignty refers to the fact that Indigenous Nations are sovereign in their governance and that extends to their data and Knowledges as well. It recognizes that Indigenous people are the ultimate authority in their data and Knowledges and aims to redefine Indigenous peoples' relationship to research from being participants or subjects to being meaningful partners and co-researchers" (UofT, 2021a).

Interpretability (or explainability): can be generally understood as the ability of humans to understand how a particular AI model works and why it generated the results that it did (Lipton, 2018; Rudin, 2019). Interpretable or explainable AI models (used interchangeably in this report) stand in contrast to **black box** models.

Open science: "inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community" (UNESCO, 2020).

Reproducibility (or replicability): extent to which consistent results are obtained when an experiment is repeated. It is one of the primary means by which the scientific community validates the accuracy of new discoveries or findings and is held as one of the "hallmarks of good science" (NASEM, 2019). It should be noted, however, that terms such as "reproducibility" and "replicability" may be defined and used in distinct and even contradictory ways by different disciplines (Fiddler & Wilcox, 2018; NASEM, 2019). Indeed, even within the field of AI research, these two terms may have different meanings depending on the source. However, this report will simply use "reproducibility" as a general term to cover a wide variety of cases and should be understood as interchangeable with "replicability."

Contents

1

Introduction

Although the potential for artificial intelligence (AI) technology to disrupt society continues to capture the imagination of the public and policy-makers, much less attention has been given to the advancements of AI in supporting design and discovery in scientific and engineering research. These advancements are nonetheless evolving rapidly, along with their impacts on both science and engineering.

AI is already beginning to drive scientific investigation by allowing for automated hypothesis generation, experiment design, experimentation, interpretation, and analysis. Technical advances and increased data availability have spurred the development and deployment of AI systems for many of these purposes in scientific research. Similar impacts may soon be felt in engineering, where the continued improvement of AI, together with advances in robotics and other related technologies, could radically alter design and manufacturing in industry. AI tools are already widely used in certain specialized settings and disciplines, and breakthroughs in AI and the application of AI in disparate fields continue to multiply. Ultimately, it is not a question of *whether* AI will eventually have an equal or larger role than humans in science and engineering design and discovery, but rather, *how quickly*?

> "Ultimately, it is not a question of *whether* AI will eventually have an equal or larger role than humans in science and engineering design and discovery, but rather, *how quickly*?"

Applications and uses of AI are diverse and follow numerous typologies. Modern AI systems are composed of hardware controlled by software that makes use of data, but these systems are written, assembled, and maintained by humans and subsequently deployed to serve a human-determined function. AI is a platform technology comprised of socio-technical systems, linked to their environments in diverse ways ranging from the training of human developers, the data these developers use, how those data were collected, and how each system evolves as a function of its environment (and potentially vice versa). Uses of AI also tend to vary according to the applications, disciplines, and data sources of interest (The Royal Society, 2019). Indeed, although practitioners in various disciplines will develop AI tools according to their own needs, progress made in discrete areas collectively serves to advance AI science overall (The Royal Society, 2019), but this can potentially complicate efforts to predict the broader impact of individual advances developed within disciplinary silos.

A legal, policy, and regulatory landscape for AI — albeit a fragmented one — is beginning to emerge despite several outstanding questions and issues. The global acceleration of AI activity has brought to light several ethical challenges in the realms of harmful biases and practices. Yet ongoing development and the opportunities provided by AI continue to be anticipated and encouraged; as this technology is introduced into society, it will be vital to ensure that it is developed concurrently with discussions on societal impact. These discussions are warranted during a technology's design and development phases rather than after its implementation.

Recognizing the opportunities, challenges, and implications of AI in science and engineering, the National Research Council of Canada (NRC), along with the Canadian Institute for Advanced Research (CIFAR), Canadian Institutes of Health Research (CIHR), Natural Sciences and Engineering Research Council (NSERC), and Social Sciences and Humanities Research Council (SSHRC) asked the Council of Canadian Academies (CCA) to convene an expert panel to provide an evidence-based and authoritative assessment of the legal/regulatory, ethical, social, and policy implications specific to the use of AI for science and engineering in Canada.

1.1 The Charge to the Panel

The CCA was asked to answer the following question and sub-questions:

What are the legal/regulatory, ethical, social, and policy challenges associated with deploying artificial intelligence technologies to enable scientific/engineering research design and discovery in Canada?

- How does the use of AI in scientific research and engineering change/influence standard science and engineering practices, processes, and outputs? Including:
    - How do we ensure scientific integrity (e.g., reproducibility, validity)?
    - What are the societal, ethical, and epistemic implications that need to be addressed?
- What are the social and ethical considerations individual researchers and research collaborations need to consider when using AI in science and engineering (e.g., bias, sex/gender)? How can current and future generations of researchers be supported to ensure ethical practices?
- What policies have been implemented related to deployment and use of AI in science and engineering in Canadian and international jurisdictions?

1.2 The Panel's Approach

In response to this request, the CCA convened a multidisciplinary and multi-sectoral expert panel (henceforth "the Panel") to address this charge, with representatives from Canada and abroad. The Panel's work was undertaken during the COVID-19 pandemic, and its process was carried out virtually as a result. The Panel met five times over the course of eight months to review evidence, discuss implications, and deliberate on its charge. As with all CCA assessments, the Panel's report was also peer reviewed prior to publication. The Panel focused on the legal/regulatory, ethical, social, and policy (LESP) challenges, including scientific integrity and epistemic and ethical issues related to deploying AI technologies to enable scientific and engineering research design and discovery in Canada.

1.2.1 Sources of Evidence

The Panel recognizes that there are numerous articles, reviews, and reports focusing on the identification of LESP issues related to AI. Although the area of focus for this report — AI for science and engineering — has been less explored, the Panel chose, when possible, to discuss the operationalization of approaches for tackling previously identified issues. The Panel wishes to note the seminal publication in April 2017 by the UK Royal Society and The Turing Institute, *Machine Learning: The Power and Promise of Computers That Learn by Example.*

A semi-structured literature review was conducted through the Web of Science library database using open-source bibliographic software (VOSViewer) to iteratively determine keywords related to recent peer reviewed literature relevant to the charge. The majority of publications drawn from the literature review were found to fall beyond the scope of science and engineering and were related to technical aspects of AI systems or LESP considerations for applications of AI that implicate society more broadly. As such, grey literature consisting of policy documents, government publications, and reports by national and international non-profit organizations constituted important sources of evidence to underpin this report.

1.2.2 Report Audiences

The issues explored in this report are relevant to a wide range of interested actors in Canada and globally. The Panel anticipates that policy-makers, decision-makers in AI industries, government agencies engaged in AI and its applications, non-profit organizations, and researchers in industry and academia may all find this report of value with respect to the risks, benefits, challenges, and considerations that accompany the development and deployment of AI for science and engineering.

1.2.3 The Panel's Interpretation of AI

The definition of AI is fluid and reflects a constantly shifting landscape marked by technological advancements and growing areas of application. Indeed, it has frequently been observed that once AI becomes capable of solving a particular problem or accomplishing a certain task, it is often no longer considered to be "real" intelligence (Haenlein & Kaplan, 2019). A firm definition was not applied for this report, given the variety of implementations described above. However, for the purposes of deliberation, the Panel chose to interpret AI as a collection of statistical and software techniques, as well as the associated data and the social context in which they evolve — this allows for a broader and more inclusive interpretation of AI technologies and forms of agency. The Panel uses the term *AI* interchangeably to describe various implementations of machine-assisted design and discovery, including those based on machine learning, deep learning, and reinforcement learning, except for specific examples where the choice of implementation is salient.

1.3 Report Scope

Early discussions with the Sponsors clarified the report's scope and goals. Of importance was how AI, applied to design and discovery, may lead to new approaches and paradigms in science and engineering that complement and potentially replace standard or conventional practices. Although the Sponsors emphasized science and engineering in their charge, social science research is very much within scope to the extent that LESP considerations are common and inherent across social science disciplines.

The scope of this report includes research and development (R&D) in the creation and use of AI tools for design and discovery across the academic, private, and government sectors. Discovery research is within the report's scope, as is commercial R&D, given the active role played by the private sector in researching, developing, and deploying AI. Basic biomedical research is also within the report's scope, given the promise shown by machine learning techniques in navigating the complex design space of research questions with biomedical relevance. However, the report does not provide a fulsome account of LESP issues surrounding AI in security or in healthcare. Other existing and ongoing work offers insight on AI's use in health research involving the direct participation of human subjects; therefore, the Panel sought to avoid this area when possible so as not to overly broaden its mandate. Following the terminology employed by the CIHR, this report prioritizes issues pertaining to pillar 1 (biomedical) research but also looks at areas that overlap with the other three pillars as the case applies.[1]

1 CIHR's four pillars of health research: biomedical, clinical, health systems services, and population health (CIHR, 2021).

1.4 Report Structure

Chapter 2 describes the context of — and motivation for — AI's application to design and discovery in science and engineering. It also provides an overview of the Canadian AI ecosystem, as well as relevant international developments related to the growth and governance of AI. **Chapter 3** provides the theoretical and foundational concepts central to deploying AI in science and engineering and the potential impact that AI would have on the nature of scientific inquiry and engineering design, such as reproducibility, interpretability, bias, ethical conduct in research, and the social practice of science. **Chapter 4** discusses the implications of using AI technologies in science and engineering research funding processes, including potential implications for the policies that govern the research system and education.

Chapters 5 and 6 examine challenges and opportunities associated with deploying AI technologies that enable design and discovery in science and engineering. **Chapter 5** identifies the social and policy implications and includes areas such as public trust, machine learning security, impacts on the labour market and environment, and managing bias in data classification. **Chapter 6** identifies legal and regulatory issues, including access to data, commercialization of innovation, liability, and security. **Chapter 7** concludes by summarizing the Panel's findings and final reflections in relation to its charge.

2

AI in a Science and Engineering Context

Chapter Findings

- Advancements in AI and its ancillary technologies will multiply opportunities for automation in science and engineering, laying the groundwork for autonomous innovation with wide-ranging implications for R&D.
- Canada's AI strategy has led to several well-developed AI research hubs providing a foundation for AI-related economic development opportunities.
- AI is emerging as a platform technology, and numerous stakeholders, policies, and trends outside of the AI environment will influence activities in the field. Awareness of these developments in adjacent areas will be necessary to effectively advance activities in science and engineering.
- Despite a rapidly evolving international regulatory landscape, efforts to establish robust guidelines for the responsible and ethical development of AI are hindered by uncertainty over governance and accountability.

The state of AI in science and engineering is undergoing rapid growth and change. Buoyed by growing volumes of data, inexpensive computing resources, and a broadening scope of application across all stages of design and discovery, AI has become a widely used tool in science and engineering. By combining algorithms and data sources, AI systems are portable and intangible compared to earlier transformative technologies and tools, which might have been fixed in place and in form. A central tension exists in finding the balance between encouraging the development of this technology — including the associated social and economic benefits — while also protecting human rights and ensuring the ethical and responsible use of AI throughout its life cycle. Indeed, decision-makers in Canada must also be mindful of developments outside of Canada because the implementation of regulatory frameworks internationally may have implications for domestic developments; this could exert pressure on policy-makers to harmonize regulations in order to avoid the disruption of international trade and commercial activities.

2.1 Harnessing AI for Design and Discovery

The original aim of AI was to take in information as a brain would and manipulate this information according to rules encoded in algorithms, arriving at decisions and outcomes more quickly and accurately than humans (Dick, 2019). Modern approaches have moved beyond systems that specialize in storing and processing vast amounts of information on specific topics, mimicking how humans would come to decisions (Anyaha, 2017). AI, mainly through machine learning systems, now encompasses a diverse collection of adaptive and dynamic techniques, many of which focus on making accurate predictions based on data (Dick, 2019), with implications for numerous forms of human cognitive labour, including the pursuit and application of scientific knowledge.

AI techniques currently used in science and engineering typically support design and discovery carried out by humans

AI has been employed in basic, applied, and experimental development research for several years, particularly in data analysis (The Turing Institute, 2021). Many of the canonical problems that can be addressed using machine learning — including classification, regression, and clustering — are highly applicable to scientific research (The Royal Society, 2017). Classification can be broadly used in image analysis, for example, while regression algorithms might facilitate the creation of predictive models based on data (The Royal Society, 2017). Digital data are created, collected, stored, and manipulated either directly or indirectly through software tools or via computer-controlled hardware. The resulting volume and diversity of data in science and engineering offer numerous insertion points into the design and discovery loop where AI systems can enable or accelerate progress beyond data analysis (Stevens *et al.*, 2020). They may be more than assistants or tools — AI systems could be deployed to autonomously pursue discovery in an exploratory bottom-up way, without being limited by cognitive biases or being driven by values or sociological constraints affecting human researchers (Kitano, 2021).

> "A central tension exists in finding the balance between encouraging the development of this technology — including the associated social and economic benefits — while also protecting human rights and ensuring the ethical and responsible use of AI throughout its life cycle."

Current uses of AI in science and engineering are diverse, and vary according to the applications, disciplines, and data sources of interest (The Royal Society, 2019). AI tools can be generative and used to produce abstract mathematical conjectures (Castelvecchi, 2021), design artificial life forms (Kriegman *et al.*, 2020), or predict the structures of vast numbers of proteins (Extance, 2021). They can also be used to simulate complex systems more efficiently than conventional tools, with implications for weather prediction (Wolchover, 2018) and computational simulations in multiple fields (Ananthaswamy, 2021a). Meanwhile, the development of AI tools for language purposes can alter the ways in which scientists interact with scientific literature due to tools that synthesize and summarize research articles (Woolston & Perkel, 2020). In commercial settings, patent databases can be searched using AI to identify "prior art" to determine the novelty and therefore patentability of an invention (Helmers *et al*., 2019). Recent extensions of language tools might eventually allow for AI systems to write computer code based on text prompts from users (Chen *et al*., 2021), with the potential to automate scientific programming tasks (Hocky & White, 2022). In these contexts, AI is mainly used as an additional tool at the disposal of scientists and engineers, augmenting or complementing conventional practices. The examples above nevertheless hint at the prospect of a more assertive role for AI in the loop of design and discovery.

Big data opens new pathways to approach design and discovery using AI, with the potential to affect the research environment and research culture

Both design and discovery involve an element of searching through possible hypotheses or lines of inquiry to unearth new knowledge or through permutations of the properties describing a system (or process) to arrive at a suitable design for accomplishing a function of interest. Design problems typically involve multiple input variables and constraints (e.g., the materials of which a component can be made or the conditions under which it operates); these combinations of variables and constraints define an enormous search space. AI continues to be harnessed by human scientists and engineers in a variety of settings to address these "combinatorially explosive" problems (The Turing Institute, 2021), where either the number of variables or the amount of data — or both — becomes unmanageably large. As the availability of data on the inputs and outputs of experiments and design problems grows, AI systems will become increasingly capable of advising human scientists on how to effectively solve a problem or reach a goal by suggesting processes they might not consider or would find counterintuitive.

For example, AI systems might propose novel layouts for experiments that permit the observation of certain phenomena in quantum physics (Ananthaswamy, 2021b), or optimize chemical reaction workflows in synthetic chemistry that defy the best practices gleaned from decades of empirical investigation (Jia *et al*., 2019). These examples illustrate how data-driven approaches can now be integrated into nearly every step of the scientific life cycle (Ezer & Whitaker, 2019) and how the role of AI systems might shift from a tool used by humans to analyze or interpret their data to one that can guide them towards novel hypotheses to explore, experiments to perform, or designs to fabricate and validate (Box 2.1).

Because big data is a key enabling factor, efforts to rapidly produce more machine-readable scientific data are multiplying. High-throughput processes in laboratories or industrial settings are also exploiting automation to generate massive amounts of data to feed into the design and discovery loop. Dedicated automated laboratories of this type are emerging to support high-throughput experimentation (NIST, 2020), and existing facilities are being updated to do so as well, due to progress in ancillary technologies (e.g., robotics, optoelectronics) (Hatfield *et al.*, 2021). In engineering and manufacturing, big data for AI systems might be created and harnessed using digital twin techniques, where data collection by sensors embedded in an object or production line allows the creation of a fully digital replica of that object or process (Boschert *et al.*, 2018). The digital twin system evolves dynamically in parallel with its physical twin, and AI systems can analyze the resulting data for monitoring purposes or to provide predictions and simulations. This has implications for the design and operation of physical infrastructure and other complex objects or industrial processes (Stevens *et al.*, 2020).

Box 2.1 Early Adoption of AI for Design in Materials Science

Some fields and disciplines, such as materials science, have been quick to identify the potential advantages conferred by data-driven techniques and the need to create and consolidate vast amounts of data (NASEM, 2018). Novel materials play a crucial role in enabling numerous technological applications, from consumer goods to infrastructure (Wang *et al.*, 2020), but it is challenging to predict how the composition of a material links to its function or how to efficiently produce arbitrary materials. Although human scientists and researchers may adjust one variable at a time to ascertain these links, AI systems might perform this search very differently (Stach *et al.*, 2021). Rather than rely solely on human intuition, experience, or trial-and-error to define which real or hypothetical material might excel at performing a given function, AI systems may be able to use correlations found in databases of material properties to allow scientists and engineers to work backwards from the desired function, potentially allowing for more rapid and efficient material designs.

Although the integration of AI into some disciplines and fields is more advanced, the growing availability of data may enable other disciplines across the natural sciences, social sciences, and humanities — some with less intuitive linkages to AI — to embrace machine learning technologies in the lab (The Royal Society, 2019). Gil *et al*. (2014) argue that academic fields have been affected by AI, many to such an extent that future progress has become almost unthinkable without the involvement of some machine learning. The high degree of international and institutional collaboration in the field of AI (Wu *et al*., 2020) suggests that the continued development of AI will accelerate the creation of new partnerships across the research environment (Chapter 4).

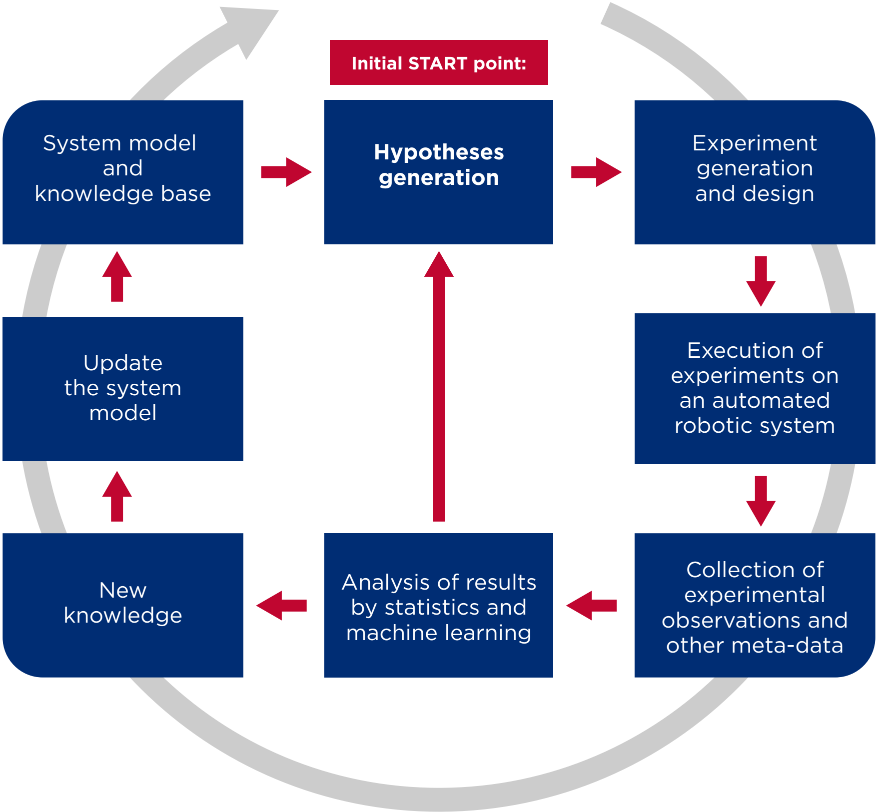
> "AI is recognized as a powerful tool for addressing interdisciplinary problems, and the development of AI benefits as much from interdisciplinary collaboration as other scientific fields benefit from the use of AI."

The U.S. National Science Foundation is creating multiple institutes to support interdisciplinary and intersectoral collaboration, with AI as a platform for addressing problems of great societal importance (Bates, 2021). Indeed, AI is recognized as a powerful tool for addressing interdisciplinary problems, and the development of AI benefits as much from interdisciplinary collaboration as other scientific fields benefit from the use of AI (Kusters *et al*., 2020). Continued advances in technology and big data are poised to alter research workflows at the scale of individual laboratories, as well as at the level of international research networks. These advances will have implications for research funding, training, and infrastructure (Chapter 4).

Autonomous design and discovery by machines with little to no input from humans may be feasible but raises several new questions and challenges

AI tools for science and engineering are widespread, but they currently lack autonomy (Gil *et al*., 2020). Although AI systems have demonstrated the capacity to support humans along the pipeline of design and discovery, significant human input and interventions have been required for the achievements attained to date by contemporary AI systems (The Turing Institute, 2021). Currently, humans are also required to extend the capabilities of an AI system from one task to another. Many AI systems are highly specialized and not easily generalized (The Turing Institute, 2021), which makes them more challenging to scale or transfer across user bases, disciplines, and sectors. These systems offer limited capacity for abstraction, and they struggle to provide scientific reasoning underpinning the discoveries made through their use[2] (Chapter 3).

2 AlphaFold, a project by DeepMind, produced a milestone 2021 result in predicting protein structures. The predictions provide input for hypotheses by human scientists, but do not explain the biophysics underlying protein folding or other biologically relevant links between structure and function (Extance, 2021).



Reproduced with permission from Springer Nature (Sparkes *et al.*, 2010)

Figure 2.1 Cycles of Automated Hypotheses Generation and Experimentation

Example workflow for an AI system capable of autonomously performing scientific research to arrive at new discoveries.

Efforts are underway to overcome these limitations and reduce the level of overt human intervention (Figure 2.1) by increasing the level of automation in laboratories, with examples in chemistry and biopharmaceutical sciences in both the academic and private sectors (Sanderson, 2019; Mullin, 2021). As human limitations are increasingly apparent in managing high volumes of data requiring rapid and accurate processing, the demand for AI solutions grows, such that an "AI fully automated system" is conceivable (King *et al.*, 2009; Sparkes *et al.*, 2010; The Turing Institute, 2021). As AI's position in the discovery cycle becomes more central, the application process for researchers is expected to change.

Several jurisdictions are deploying programs that aim to realize fully autonomous AI systems for science and engineering applications in the context of Grand Challenges. The Government of Japan is funding a cross-cutting initiative to produce robotic systems powered by AI — ones that are capable of realizing "impactful scientific principles and solutions" among other societally beneficial goals (CSTI, 2020). Elsewhere, the Turing Nobel Challenge, launched in the United Kingdom but framed around international and interdisciplinary collaboration, targets the creation of an "AI scientist" capable of making discoveries worthy of a Nobel Prize by the year 2050 (Gil *et al.*, 2020; Kitano, 2021). The resulting system — effectively an autonomous researcher — should be capable of demonstrating expert knowledge, communicating results and discoveries to humans, and generating compelling research questions (Gil *et al.*, 2020). The system would formulate hypotheses, as well as the necessary experiments (Kitano, 2021), and could determine the priority of exploration using various means, including by estimating potential impacts (Chapter 4).

> "Rather than replace human scientists outright, AI systems could be harnessed to offer a novel method of scientific discovery; still, many ramifications remain uncertain."

Rather than replace human scientists outright, AI systems could be harnessed to offer a novel method of scientific discovery; still, many ramifications remain uncertain. From a social standpoint, AI systems of this nature would displace human labour to an unknown degree and would be deployed in social contexts far different from those in which they were developed (Chapter 5). For example, concerns surrounding this issue have led to calls for AI systems to be co-developed by researchers and Indigenous communities for applications where deployment would occur in those communities (Lewis, 2020) (Section 5.2).

Other potential consequences are more ambiguous and uncertain, including how access to AI resources could be allocated in a way that avoids exacerbating pre-existing inequalities in the research system. An autonomous researcher could foreseeably break path-dependency in research because it would be stripped of certain problematic incentives or values that influence research directions made by humans (Kitano, 2021). However, it would still be desirable for such an AI scientist to operate in a manner aligned with human values (The Turing Institute, 2021); after all, the research dedicated to the development of this autonomous scientist will invariably require significant investments by both public and private sectors. Yet, as with autonomous vehicles or AI-driven recommender systems, explicitly integrating human values into algorithmic systems is not always feasible or practical (Gibert, 2020), nor are decisions about which values to integrate straightforward.

Finally, although it remains a challenge to encode domain expertise, training, and human experience into AI systems (King *et al.*, 2018), the reverse is also a concern. For humans to effectively use the knowledge created by AI, the discoveries should be interpretable, timely, and relevant to human scientists (The Turing Institute, 2021). The epistemic and ethical challenges that are explored more deeply in subsequent chapters represent significant hurdles to overcome along the path towards design and discovery with humans outside the loop.

2.2 Canadian Context

CIFAR established long-term research programs in AI from the 1980s onwards, supporting foundational research at the field's nadir in popularity (Johnson, 2021). At that time, researchers exploring deep learning and reinforcement learning were considered to be on the periphery of the computer science community. However, these approaches are now widely understood as significant implementations of AI with Canadian origins (Smith, 2017). In the past decade, competition in AI development has accelerated on a global scale as the field increasingly moves beyond academic research environments and into society. In developing AI systems for science and engineering, Canada will be faced with managing its intrinsic strengths and weaknesses in R&D in an increasingly multipolar and dynamic AI landscape.

The current AI landscape in Canada was seeded by long-term public investments and is presently dominated by a small number of regional actors that drive continued growth

Canada's competitiveness in AI can be traced back to consistent and early investments in the field, even during times when it was unfashionable. Several key figures in AI research established their careers in Canada, in part due to stable public funding in support of AI researchers and research activities (Gherhes *et al.*, 2021). More recently, the 2017 launch of the Pan-Canadian AI Strategy marked a new major investment of public funds that has shaped the present landscape for Canadian AI. The Pan-Canadian AI Strategy, led by CIFAR, represented the first comprehensive national strategy of its kind (UNESCO, 2018), providing a timely stimulus to take advantage of the groundwork laid during previous decades (Smith, 2017). The strategy focuses on supporting activities at three main centres for AI research: the Alberta Machine Intelligence Institute (Amii), the Vector Institute, and the Quebec Artificial Intelligence Institute (Mila) (Smith, 2017). These three non-profit institutes represent important hubs in Canadian AI R&D by hosting research chairs and scholars and by sustaining networks among higher education institutions and industry, including numerous multinational firms active in AI R&D (Chowdhury *et al.*, 2020). British Columbia has also developed strengths in AI, specifically in machine vision, and benefits from a strong start-up ecosystem and a favourable environment for the commercialization of AI applications (Chowdhury *et al.*, 2020).

Other specialized initiatives focusing on the application of AI to problems in science or engineering continue to appear. The NRC AI for Design Challenge program develops and provides AI technologies and capabilities to accelerate discovery, R&D, and innovation processes. It advances algorithms, methods, and datasets to assist engineers, researchers, and scientists with design and scientific discovery (NRC, 2021). The federal Innovation Superclusters Initiative, meanwhile, was launched in 2018 with the aim of creating centralized hubs of R&D activity across Canada in numerous fields to address shortcomings in innovation and technology (Knubley, 2021). Located in Montréal, the Scale AI Supercluster focuses on applying AI to commercial products and supply chains, but other superclusters are also active in AI applications or other relevant peripheral sectors such as robotics (Knubley, 2021). In 2021, the University of Toronto launched the Acceleration Consortium for materials research which includes more than 50 top investigators from the university and internationally (Kalvapalle, 2021). The consortium is largely focused on identifying and producing promising new advanced materials for technology development; however, it will also be oriented towards capacity building and research into laboratory automation in general, AI-aided experimental design, and training of highly qualified personnel in this field (Acceleration Consortium, 2021).

Canada's AI research strategy has created momentum and established critical mass in its hubs, but enabling design and discovery in science and engineering will require expanding the scope of current activities

The Canadian AI research environment is specialized and possesses strengths in several fields, including deep learning, computer vision, and reinforcement learning (Chowdhury *et al.*, 2020).[3] Early impressions suggest that the Pan-Canadian AI Strategy accomplished several of its aims, and the 2021 federal budget committed new funding over the coming decade for the strategy's renewal (Chowdhury *et al.*, 2020; GC, 2021c). The contributions and impacts of AI researchers in Canada are reflected in relatively high rankings in traditional indicators linked to publication and patenting activity (OECD, 2021e). Canada also benefits from a high rank in the Global AI Index (Tortoise Media, 2021), a figure based on a combination of several indicators weighted to provide an aggregate overview of a country's AI capacity across multiple sectors (Tortoise Media, 2020). Canada's world-leading research capacity in some sub-fields is an important reflection of long-term and focused government support for fundamental, investigator-led AI research (Gherhes *et al.*, 2021).

> "In the Panel's view, the Canadian AI ecosystem — thus far accustomed to vertical growth — will need to grow horizontally beyond its existing strengths, crossing physical, disciplinary, and sectoral boundaries to make the most of opportunities for design and discovery in science and engineering."

Nevertheless, some elements of the current landscape may undermine Canadian competitiveness with respect to AI applications in science and engineering. The availability of training and educational programs in AI-related fields in Canada skews heavily towards specific areas, namely robotics and automation, whereas other jurisdictions offer training programs in a broader selection of AI sub-fields (OECD, 2021b). Canada was early to define a national strategy, but that strategy focused heavily on research and talent in AI specifically, whereas more recent international examples provide greater consideration to actors beyond the research ecosystem (Kung *et al.*, 2020). Concerns have also been raised that the approach thus far of consolidating funding and activities to a handful of entities and geographic areas could be problematic (Brandusescu, 2021). While evidence suggests that increasing funding concentration beyond a certain threshold leads to diminishing marginal returns from research investments (Fortin & Currie, 2013; Lorsch, 2015; Mongeon *et al.*, 2016), a key architect of the supercluster program asserts that consolidation

3 Patenting activity also demonstrates a degree of specialization, particularly in fields like natural language processing (ISED, 2019a).

addresses concerns expressed by some stakeholders that Canadian resources in science, technology, and innovation are spread too thin (Knubley, 2021).

Such tensions are inevitable, and although many applications will benefit from the strengths of local networks, AI is an inherently portable platform. Hubs will therefore need to be well connected to geographically dispersed activities across Canada in order to effectively leverage national R&D capacity. In the Panel's view, the Canadian AI ecosystem — thus far accustomed to vertical growth — will need to grow horizontally beyond its existing strengths, crossing physical, disciplinary, and sectoral boundaries to make the most of opportunities for design and discovery in science and engineering.

2.3 Developments in Governance and Policy

Despite rapid growth in the field, the use of AI as a platform technology is not currently subject to comprehensive regulation. The potential for AI to cause harm, exacerbate inequities, or be misleading, among other risks, has prompted multiple attempts to define guidelines for its ethical development. Many aspects of governance, nevertheless, remain uncertain, from substantive issues and differences reflected in international declarations to roles and responsibilities for implementing and enforcing governance principles. Different approaches are beginning to take shape across the spectrum, from "hard" regulation via legislation to "soft" regimes, where practitioners commit to voluntary standards and codes of practice. Some AI applications for science and engineering are not inherently controversial; for example, the identification of objects or anomalies in astronomical data does not impact human rights. Nonetheless, and particularly as AI systems become more autonomous, the developments in AI governance described below — as well as additional trends in data governance and research culture — will inevitably come to define the space in which researchers and innovators create, collaborate, and practise.

The multiple national and international guidelines published to date on AI ethics and governance are voluntary and adhered to inconsistently

Many sets of guidelines have been proposed by governmental and non-governmental organizations around the world — frequently following multistakeholder consultations — to promote responsible, human-centred AI developments. The scope of these guidelines is broad, and the diversity of their proposed recommendations highlights tensions that exist between the promotion of innovation and the need to protect human rights or between local and international considerations, among others (Gibert *et al.*, 2018). These tensions arise due to the large number of stakeholders in AI who may possess competing or contradictory priorities. For example, the Montréal Declaration for a Responsible Development of Artificial Intelligence represents a Canadian-based initiative in AI governance with numerous signatories in Canada and abroad, including public and private organizations, as well as individuals (Montréal Declaration for Responsible AI, 2021). In contrast, the Toronto Declaration provides guidelines for preventing the development of discriminatory machine learning systems on the basis of international human rights laws (Amnesty International & Access Now, 2018). At the national level, as of October 2021, the governments of 46 countries have officially signed on to the principles outlined by the OECD's Council on AI. Although these guidelines provide a set of internationally agreed-upon principles, they are not legally binding, and the vast majority of signatories are themselves OECD member countries (38 of the 46) (OECD, 2021c). The uptake or adherence to any one set of guidelines is variable and follows from the priorities of individual stakeholders or governments.

Common themes can be found within existing guidance documents. They consistently emphasize well-being, autonomy, justice, privacy, knowledge, democracy, and responsibility while proposing frameworks that span the spectrum from soft to hard regulation (Gibert *et al.*, 2018). A review of 84 ethics guidelines for AI identified a subset of principles, such as transparency, justice and fairness, non-maleficence, responsibility, and privacy, to be most prominent. The analysis found variability in the prioritization and interpretation of these principles, as well as a lack of clarity regarding to which stakeholders the guidelines should (or should not) apply (Jobin *et al.*, 2019). Although there is no clear consensus on universal guidelines to govern AI, more recent documents have begun to converge on similar themes (Fjeld *et al.*, 2020). Nevertheless, concerns persist that the voices of certain key stakeholders remain unheard. Although many guidelines were developed through participatory mechanisms in the hopes of capturing input from the broad diversity of stakeholders who will be impacted by AI, individuals or organizations from liberal Western democracies tend to be overrepresented (Gibert *et al.*, 2018). Jobin *et al.* (2019) found that, at a global level, Africa, South and Central America, and Central Asia were under-represented in their review of ethical guidelines. Achieving greater democratization and more inclusive participation in AI may allow for new priorities to be added to existing principles (Fjeld *et al.*, 2020).

The path from AI ethics guidelines to actions is not specified in most documents, nor are frameworks for resolving conflicts among principles or for ensuring compliance described in these guidelines (Jobin *et al.*, 2019; Whittlestone *et al.*, 2019). As such, many gaps exist in the implementation of declarations on ethical or responsible AI (Whittlestone *et al.*, 2019; Fjeld *et al.*, 2020), and soft approaches to regulation are presently the global norm (OECD, 2021b). Individuals and organizations actively applying AI to science and engineering will need to remain mindful of the fluid status of AI governance at international, national, and sub-national scales. Morley (2021) attempts to close the gap between principles and practices by establishing a typology to support developers in applying ethics at each stage of the machine learning development pipeline. Although their work is focused solely on machine learning, the research may be applicable to other AI categories. See Section 3.2 for additional discussion of the application of ethical principles in the use of AI for science and engineering.

AI applications will increase linkages and interdependencies among the AI community and the wider science and engineering landscape, with implications for policy and governance

AI systems are not limited to the algorithms that define their operation. These socio-technical systems also rely on data, complementary technology (e.g., sensors), and the social contexts in which they are developed and deployed (OECD, 2019b). Efforts and strategies to promote AI must therefore be mindful of intersections among policy developments occurring in adjacent areas.

In the academic community, changes to incentive and reward structures in scientific research impact the trajectory of AI for science and engineering, for example through the allocation of research funding. Research funding agencies continue to evolve methodologies for the assessment of research and researchers (GRC, 2021), with numerous implications if humans are taken out of the loop (Chapter 4). For instance, the European Union's (EU) Responsible Research and Innovation (RRI) initiative seeks to effect changes in research culture by promoting approaches to science and innovation that are ethical, inclusive, and open (RRITools, n.d.).

Additionally, the OECD has recommended supporting open data to promote access to data across all sectors as a way to facilitate the development of AI (OECD, 2019a). Therefore, investments in data infrastructure and developments in data policy are also instrumental in promoting the uptake and adoption of AI. These investments might include the Internet of Things, broadband access, and high-performance computing and storage (OECD, 2021b). The evolution of data policy frameworks will influence activities in AI for science and engineering, even if these frameworks do not target innovation in those areas as a specific focus, such as the United Kingdom National Data Strategy (Gov. of the UK, 2020). See Chapter 6 for further discussion on access to data.

The hopes of using AI to address other global policy goals will also influence the pace of AI development and the prioritization of applications, particularly in science and engineering.

3

Implications of AI for the Conduct of Research

Chapter Findings

- The field of AI research currently has problems with reproducibility, which could complicate its use in other scientific disciplines.
- Some popular types of AI systems operate as black boxes, such that it may be difficult or even impossible to explain how their results were generated; this can hinder scientific explanation and understanding and potentially undermine the credibility of machine-generated scientific findings.
- The accuracy of results produced by AI is determined, in part, by the quality of the data on which it has been trained. AI systems may produce inaccurate or skewed results due to biases in training datasets and problems of generalizing from training data to new data.
- Current ethical frameworks for the conduct of research do not address the complexities that AI brings to traditional notions in research ethics, such as that of human research participants and informed consent.
- The increased use of AI could affect the relative importance of different values in science and engineering research, as well as their underlying social dynamics.

The increased use of AI for science and engineering research is likely to affect the nature and practice of scientific inquiry, including researchers' understanding of scientific integrity, ethical conduct in research, and the values and social practices of science. Interpretable and explainable AI will be essential in the context of science and engineering, to enable human scientists to learn from the discoveries and designs of AI, mitigate risks, and promote trust and transparency in research. New ethical and responsible frameworks for research utilizing AI may also be needed to address novel challenges relating to issues such as informed consent, social harms arising from AI, and the use and reuse of data. Moreover, in addition to epistemic and ethical values, AI may also shift the values that govern the social practice of science, such as those underlying the validation, provenance, and dissemination of scientific results.

3.1 Ensuring Scientific Integrity in Research that uses AI

The increased use of AI in scientific and engineering research brings new epistemic and ethical challenges for ensuring scientific integrity in research. Scientific integrity covers a wide variety of issues, including the management of conflicts of interest, acknowledgement of contributors and authorship, and use of research funds, among others (CCA, 2010). This section focuses on features of scientific integrity that give rise to unique challenges in the context of AI for science and engineering: reproducibility, interpretability, accuracy and bias, and data management. Specifically, findings generated by AI may be difficult to validate because of barriers to independent reproducibility, as well as a lack of explainability or interpretability of some types of AI systems and their results. Consequently, it may be difficult to verify the accuracy of the discoveries generated by AI and ensure that they are free from undue bias. Addressing these issues will require transparency and accountability in research practices and in data stewardship and management.

3.1.1 Reproducibility

Reproducibility is one means by which the scientific community validates the accuracy of new discoveries or findings and is held as one of the "hallmarks of good science" (NASEM, 2019). Broadly speaking, reproducibility refers to the ability of independent researchers to achieve the same (or similar) results as a previous study using the same (or similar) methods, thereby demonstrating that study's validity.[4] Although AI has the potential to improve reproducibility in science (King *et al.*, 2009), its incorporation into the scientific process may also create new challenges (Carter *et al.*, 2019; The Royal Society, 2019). Some argue that the field of AI research has a significant reproducibility problem (Hutson, 2018; Heaven, 2020), which could negatively affect other fields of scientific and engineering research that use AI.

AI has the potential to improve reproducibility in science

AI has the potential to improve reproducibility in science by enabling the description and recording of experiments in greater detail and semantic clarity than humans can (King *et al.*, 2009, 2018). Because information about experimental setup and procedures — i.e., "experimental metadata" (King *et al.*, 2009) — is automatically recorded by AI systems in perfect fidelity using formal

4 Terms such as "reproducibility" and "replicability" may be defined and used in distinct and even contradictory ways by different disciplines (Fiddler & Wilcox, 2018; NASEM, 2019). Indeed, even within the field of AI research, these two terms may have opposite meanings depending on the source; see, e.g., Gundersen and Kjensmo (2018) and Nagarajan *et al.* (2019) versus NASEM (2020) and Carter *et al.* (2019). However, this report will simply use "reproducibility" as a general term to cover a wide variety of cases and should be understood as interchangeable with "replicability."

logical languages, AI-powered autonomous researchers (or "robot scientists") can perfectly recreate all relevant aspects of an experiment. By contrast, not only can humans make errors or omissions when recording such information, but such activities are also time-consuming and require knowledge of reporting standards. Moreover, unlike AI, humans typically record experimental metadata in natural languages that can introduce vagueness and imprecision, which inhibits reproducibility (King *et al.*, 2018).

Increased transparency is needed to facilitate reproducibility in AI research

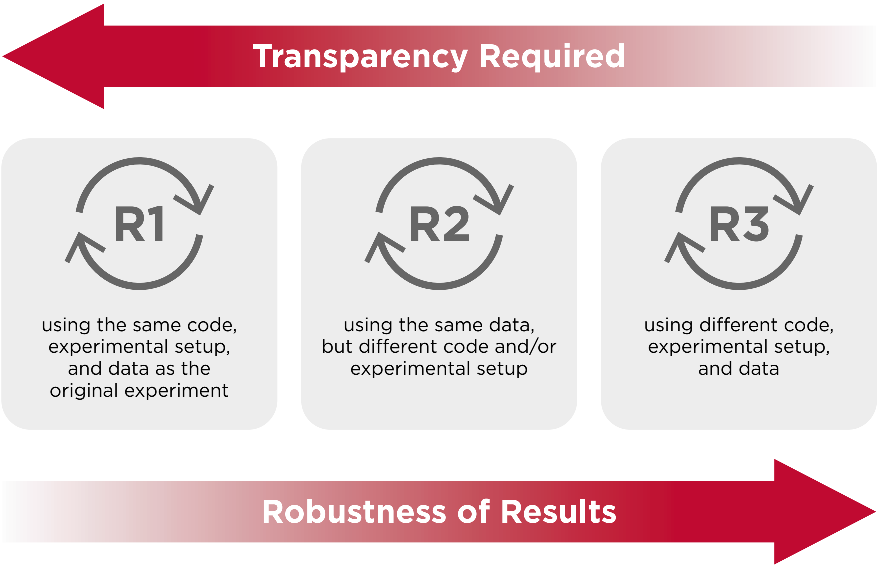
One of the main challenges to scientific reproducibility raised by AI is a lack of transparency on behalf of researchers, who may not provide sufficiently complete or detailed information to allow others to replicate their findings. In order for machine-generated results to be reproducible, researchers need to provide information about the code, data, and computing infrastructure on which the experiment was performed (The Royal Society, 2019; Haibe-Kains *et al.*, 2020; Heaven, 2020). Yet a literature review by Gundersen *et al.* (2018) found that only about 8% of papers presented at top AI conferences between 2013 and 2016 shared their code, and only about one-third shared their test dataset(s).

> "Consequently, it may be difficult to verify the accuracy of the discoveries generated by AI and ensure that they are free from undue bias. Addressing these issues will require transparency and accountability in research practices and in data stewardship and management."

Merely providing code is often insufficient for reproducibility in AI research because the same code can produce different results when executed under different experimental conditions, such as hardware or compilers, software environments, hyperparameters (i.e., variables that control the learning process, such as the architecture of a neural network or the learning rate), or random seeds (i.e., factors used to initialize the weightings of connections in a neural network) (Henderson *et al.*, 2018; Hutson, 2018; Nagarajan *et al.*, 2019). Thus, reproducibility generally requires researchers to provide additional information about the specific conditions under which their models are trained and tuned. Providing access to the data from which the AI models were derived is also key to reproducibility (Haibe-Kains *et al.*, 2020) because the same code can provide different results when trained on different datasets (Hutson, 2018). Relevant data to share include training data, validation data, test data, and results (Gundersen *et al.*, 2018). Barriers to transparency in data sharing include protection of proprietary information, licensing issues, or privacy concerns (in cases where the data involve human subjects) (Haibe-Kains *et al.*, 2020; Heaven, 2020).

Reproducibility in AI research exists on a spectrum

It is possible to distinguish between different degrees of reproducibility in AI research. Broadly speaking, researchers may attempt to reproduce the same (or similar) results as the original experiment using: (R1) the same code, experimental setup, and data; (R2) the same data, but different code or experimental setup; or (R3) different code, experimental setup, and data (Gundersen *et al.*, 2018) (Figure 3.1). Notably, R1 demands the most transparency on behalf of AI researchers because it requires precise documentation of (or access to) the original code, experimental setup, and dataset(s); by contrast, R3 demands the least amount of transparency because it only requires a sufficiently detailed description of the experimental goals and method. However, the results of R1 are also the least robust or generalizable, as they only confirm they can be replicated under highly specific, narrow conditions. R3 is the most robust and demonstrates that the same results can be generated using different approaches and different data (Gundersen *et al.*, 2018).



Adapted with permission: Copyright © 2018, Association for the Advancement of Artificial Intelligence (Gundersen *et al.* 2018)

Figure 3.1 Degrees of Reproducibility in AI Research

Reproducibility may take different forms and serve different functions across disciplines

It is important to differentiate between reproducibility in *AI research* and reproducibility in other fields of scientific research that *use AI*. Reproducibility can take different forms in different scientific disciplines and can have different goals and serve different types of epistemic functions (Leonelli, 2018). For example, attempts to reproduce research involving model organisms in experimental psychology and neuroscience may not aim to directly reproduce previous results but may instead seek to identify sources of variation across experimental setups that can alter the interpretation of the data produced by the experiment. Research involving rare, perishable, or inaccessible objects or materials — such as unique organic specimens or archaeological or paleontological remains — often cannot be directly reproduced because the uniqueness and irreproducibility of the object or materials are central to their evidentiary value. Research in medicine, history, and the social sciences often relies on observation rather than experimentation, wherein attempts at reproducibility rely on the expertise of skilled observers (Leonelli, 2018). In short, because the goals and function of reproducibility vary across scientific disciplines, they may correspondingly vary depending on the disciplines in which AI is being used. As such, it may not be appropriate or useful to apply a single set of reproducibility standards to both AI research and other fields of research that use AI.

Several initiatives exist to facilitate better reproducibility in research using AI

There are numerous online platforms for making AI research more transparent and reproducible, such as GitHub to share code and TensorFlow to share AI models and frameworks (Haibe-Kains *et al.*, 2020). Isdahl and Gundersen (2019) provide a comparative analysis of which online machine learning platforms — such as OpenML, BEAT, or Floydhub — best support reproducibility and in which ways. In cases where a dataset cannot be shared, alternatives include providing information about data-collection techniques or data labels so that others can build similar datasets, or allowing independent auditors access to the data to verify the results (Haibe-Kains *et al.*, 2020). Gebru *et al.* (2021) have proposed *datasheets for datasets*, which could help facilitate reproducibility in machine learning research by providing detailed metadata about datasets. This includes the motivation for their creation, their composition, their collection process, the preprocessing and labelling process, their recommended uses (including distribution and maintenance), and any other relevant features (Gebru *et al.*, 2021). To enhance replicability in AI research, engineers at IBM Research developed an AI tool that recreates unpublished source code based on descriptions in a paper (Sethi *et al.*, 2018).

Carter *et al.* (2019) have proposed several potential solutions to the challenges of sharing code and data. These include (i) a shareable, video-based "virtual review," roughly equivalent to a regulatory inspection, in which code is discussed in detail and demonstrated on-screen; (ii) creating a "protected computing environment," in which the data and modelling would be available in a read-only, non-exportable format for reviewers; and (iii) making the data, code, and computing infrastructure available to reviewers through appropriate licensing agreements that legally ensure data security, confidentiality, and intellectual property (IP) protection.

Academic publishers, editors, and conferences also have a role to play in fostering reproducibility. For example, the AI conference NeurIPS recently developed a checklist of disclosure requirements that researchers must provide when they submit papers in order to help increase transparency and reproducibility (NeurIPS, 2021). Similarly, the JOURNE workshop, held in the spring of 2021 as part of the Conference on Machine Learning and Systems (MLSys), focused on increasing transparency in machine learning research by providing critical reflections on negative results, intermediate findings, and the process of developing research ideas in order to help correct biases towards publishing only "good results" (JOURNE, 2021). Initiatives such as these are important to help facilitate reproducibility because current research and publishing incentives (in AI, but also other fields) are generally not conducive to reproducibility due to the pressure to publish quickly, the lack of time to test algorithms under different conditions, the lack of space to document all attempted configurations of the experimental setup, and the disincentive to publish failed attempts at replication (Hutson, 2018; NASEM, 2019).

3.1.2 Interpretability, Explainability, and the Black Box Problem

One of the most promising features of AI is its ability to identify patterns in data that are invisible to humans, enabling new scientific discoveries (Samek & Müller, 2019; The Royal Society, 2019). However, several popular AI methods — such as those based on deep learning neural networks — are often able to produce highly accurate results, yet may still operate as black boxes, such that their users and even designers may be unable to explain how the results were generated or upon what features in the data the results are based (Knight, 2017; The Royal Society, 2019). Although, in some cases, accuracy alone may be sufficient for scientific progress, the goal of science is ultimately *explanation* and *understanding* (The Royal Society, 2019). As Samek and Müller (2019) put it:

> *In the sciences [...] explaining and interpreting what features the AI system uses for predicting, is often more important than the prediction itself, because it unveils information about the biological, chemical, or neural mechanisms and may lead to new scientific insights.*

For these reasons, researchers in some scientific fields have come to prefer AI models that are interpretable and explainable, even at the expense of predictive accuracy (Lapuschkin *et al.*, 2019). Indeed, there may sometimes be a tension between explainability and predictive accuracy; some of the most predictively accurate AI systems are among the most opaque and least explainable, whereas the most transparent and explainable AI systems may have lower levels of accuracy (Gunning & Aha, 2017). However, the idea that there is a necessary trade-off between interpretability and accuracy has been challenged by Rudin and Radin (2019), who review examples of interpretable AI systems that achieve accuracy equivalent to (or better than) black box AI systems in domains such as criminal justice, healthcare, and computer vision.

There are many reasons to prefer interpretable models over black boxes in the context of AI for science and engineering. In black box AI systems, it may be more difficult to identify cases where an AI system is making correct predictions but is doing so based on spurious correlations (Samek & Müller, 2019). For example, a black box AI system was able to accurately classify images in a widely used benchmark dataset of objects such as boats, airplanes, and horses, yet a subsequent analysis discovered that it achieved its predictive accuracy, in part, based on the frequent presence of source tag watermarks on images of horses and a particular kind of patterned border on images of airplanes (Lapuschkin *et al.*, 2019). If an AI system is a black box, it may be next to impossible to identify such spurious correlations. Science typically aims to understand causal relationships in the natural world, yet the associations learned by AI do not necessarily reflect causal relationships (Lipton, 2018). Furthermore, the ability to interpret the results of an AI model allows for an iterative process of refinement and processing to improve the model's accuracy, thereby potentially leading to better accuracy overall compared to black box models (Rudin & Radin, 2019). The interpretability of AI models is important for detecting potentially unfair or discriminatory outcomes (Section 5.2), as well as for ensuring data privacy (Lipton, 2018; Rudin, 2019).

"Some have suggested that the complexity and inherent opacity of certain types of AI systems raise the question of whether scientific results produced by AI may be (or could become) unintelligible to human beings."

Some have suggested that the complexity and inherent opacity of certain types of AI systems raise the question of whether scientific results produced by AI may be (or could become) unintelligible to human beings (Nickles, 2018). That is, the findings produced by AI may be what Humphreys (2009) calls "epistemically opaque," such that humans may lack the cognitive capacity to fully comprehend the process by which an AI system produces those findings (Leonelli, 2020). This is related to the concept of *ultra-strong machine learning*, which refers to the ability of an AI to generate a hypothesis that (i) it can teach to a human, and (ii) consequently allows the human to increase their predictive performance on a task beyond that of a human who studies the training data alone (Muggleton *et al*., 2018). As such, ultra-strong machine learning can be understood as a measure of human comprehension and cognitive ability relative to AI and the transferability of novel AI reasoning capabilities to humans.

Explainable AI (XAI) aims to create transparent AI systems that are more understandable to humans

In recent years, the concept of XAI has gained a great deal of attention, with several workshops and initiatives dedicated to it (Adadi & Berrada, 2018). For example, in 2017, the U.S. Defense Advanced Research Projects Agency (DARPA) launched its XAI program with the goal of creating new or modified machine learning techniques that enable explainability and understanding (Gunning & Aha, 2017). Although there is no standard or generally accepted definition of what constitutes XAI (as the term refers more to a movement than to a specific concept), it is widely acknowledged that XAI is necessary for trust, understanding, and effective management of AI results because it allows users to understand an AI system's strengths and weaknesses and permits the identification of potential errors or bias (Gunning & Aha, 2017; Adadi & Berrada, 2018).

There are a variety of methods and techniques used to make AI models more interpretable (see Adadi and Berrada (2018), Samek and Müller (2019), and Lipton (2018)). These techniques differ with regard to whether the models are intrinsically interpretable or only interpretable by post-hoc explanation; whether the interpretations they provide are global (explaining the logic of the entire model) or local (explaining only a particular result generated by the model); and whether they are specific to certain types of AI models or apply to any class of AI model (Adadi & Berrada, 2018).

3.1.3 Accuracy and Bias

Ensuring that results generated by AI are accurate and unbiased is vital to its use in science and engineering. AI techniques can be susceptible to the problem of *overfitting*, in which a machine learning algorithm is able to produce accurate results when applied to its training data but produces inaccurate results when applied to new data (Ying, 2019). This problem may be due to several factors, such as a small-sized dataset, a low signal-to-noise ratio in the dataset (i.e., when there is a significant number of spurious correlations in the dataset that make it difficult to identify meaningful correlations (Lehr & Ohm, 2017)), or simply a poorly calibrated algorithm. There are a variety of solutions to this problem, which range from improving or expanding the training data to selecting different models to optimizing model hyperparameters. Other solutions include *ensembling*, in which predictions for differently trained algorithms are combined to produce a more accurate result, and *cross-validation*, in which the same algorithm is applied to different subsets of the data to enable comparison. In the case of AI systems based on neural networks, developers may make use of technique-specific approaches such as *dropout*, in which nodes in the network are randomly dropped during training (see Bejani and Ghatee (2021) and Ying (2019) for a review of different solutions to overfitting).

> “It is crucial to consider the forms of bias associated with AI systems in tandem with the systems of accountability associated with decisions around data governance, management, and processing.”

The accuracy of the results produced by an AI algorithm is determined (in part) by the quality of the data on which it has been trained. Large datasets that have not been subject to close scrutiny and quality checks are at a higher risk of containing inaccuracies and unaccounted-for biases that could result in misleading or incorrect results (Leonelli, 2020). Moreover, human biases in constructing datasets for scientific research can affect AI models used to advance that research. For example, Jia *et al.* (2019) demonstrated that an AI model trained on a smaller dataset of randomly selected chemical reactions outperformed a model trained on a larger, human-selected dataset in predicting reaction outcomes. This is because the human-selected dataset was influenced by factors such as the relative popularity of certain reactants or reaction conditions and the frequency with which they appear in scientific publications (Jia *et al.*, 2019).

This example demonstrates how biases in AI models can result from decisions around the collection and curation of training data. This is not an entirely neutral or objective process because researchers must make decisions about what data will be collected (and how) and what will be rejected or ignored (boyd & Crawford, 2012). Furthermore, biases in datasets may result from practical considerations that are not purely epistemic, such as those related to data packaging, storage, and sharing (Leonelli, 2020). Researchers must also make a variety of subjective decisions, including the choice of target variables, whether and how to label training data, how to treat outliers, how to partition data for testing, which algorithms to choose, and how to tune the model (Lehr & Ohm, 2017; Selbst & Barocas, 2018) — all of which can introduce additional bias.

It is crucial to consider the forms of bias associated with AI systems in tandem with the systems of accountability associated with decisions around data governance, management, and processing. In the view of the Panel, retaining oversight over who makes such decisions — and how these decisions are made — is imperative to the critical scrutiny of AI systems and their impacts, both in terms of the reliability of the knowledge produced and of its ethical value. Documentation of the choices made during the AI model development process can help to facilitate such oversight (Selbst & Barocas, 2018).

In the context of AI, bias can mean several different things

The term *bias* is often used in the context of AI; however, it can have different meanings and be understood in different ways, which can sometimes be conflated (Hellström *et al.*, 2020). In common usage and in many media reports about AI, the term is often used to describe AI models that lead to discriminatory outcomes — for example, "when unfair judgments are made because the individual making the judgment is influenced by a characteristic that is *actually* irrelevant to the matter at hand, typically a discriminatory preconception about members of a group" (Muller, 2020). These types of biases in AI raise social and ethical issues, which are discussed further in Chapter 5. Importantly, however, this type of social and ethical bias should be distinguished from *statistical* bias, which can result from a variety of factors, such as sampling or measurement errors.

Finally, *bias* has another established meaning in AI, referring to the types of biases that are necessary for an AI to function. For example, to make inductive inferences that allow a machine learning system to generalize from its training data to new examples, such a system requires some form of built-in bias based on assumptions about the data source. This is known as *inductive bias* (Amit & Meir, 2019; Hellström *et al.*, 2020). Similarly, designers of AI systems must make choices about the statistical threshold at which a classification is accepted (*uncertainty bias*), as well as about the choice of features that constitute inputs and outputs for the model (*specification bias*) (Hellström *et al.*, 2020).

3.1.4 Data Stewardship and Management

Ensuring scientific integrity in research utilizing AI requires the implementation of data stewardship and management principles that facilitate responsible and ethical data sharing and use. The FAIR (Findable, Accessible, Interoperable, Reusable) data principles, which aim to facilitate the reuse of scientific data (Wilkinson *et al.*, 2016), have become a widely adopted standard for data management and stewardship in science (Mons *et al.*, 2017; Boeckhout *et al.*, 2018). The U.S. Department of Energy has noted the importance of FAIR data principles for AI in science and stated that "AI systems can be employed to automate the creation of FAIR data and integrate it into knowledge repositories, in turn providing the architectural basis for new data infrastructure necessary to accelerate AI training and model development" (Stevens *et al.*, 2020).

Socially and ethically responsible data stewardship and management for AI in science and engineering may require additional measures beyond compliance with the FAIR principles

Although the FAIR principles focus on practical issues related to data sharing and distribution, they do not deal with social and ethical issues associated with data stewardship and management. Consequently, other data management principles have been proposed to complement FAIR, including:

- the TRUST (Transparency, Responsibility, User focus, Sustainability, Technology) principles for digital repository trustworthiness aim to provide a framework to "facilitate discussion and implementation" of data preservation and archiving (Lin *et al.*, 2020).
- the FACT (Fairness, Accuracy, Confidentiality, Transparency) principles, developed by the Responsible Data Initiative, aim to facilitate ethics and responsibility in data science by focusing on foundational scientific challenges (van der Aalst *et al.*, 2017).
- the CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles for Indigenous data governance, developed by the Indigenous Data Sovereignty Group (within the Research Data Alliance (RDA)) as a complement to the FAIR principles, highlight the role of data in advancing self-determination and innovation among Indigenous Peoples (Carroll *et al.*, 2020) (Section 5.2.1).

In addition, some organizations are working to address the issue of data sharing; for example, the RDA is an international organization that aims to develop and disseminate the technical, social, and community infrastructure needed to facilitate data sharing and data-driven exploration, particularly in the academic research community (Berman, 2019). Within the RDA, researchers from different scientific disciplines and fields of research develop Interest Groups that focus on the unique infrastructure needs — including code, protocols, tools, models, curricula, policies, and standards — that support data-based research in their disciplines (Berman, 2019).

3.2 Ethical Use of AI in Science and Engineering Research

The ethical use of AI has received a great deal of attention, particularly in areas such as decision-making (both by governments and private organizations), law enforcement, facial recognition, online privacy, misinformation and "deepfakes," manipulation of behaviour, and autonomous systems such as vehicles and weapons (Muller, 2020). Less explored, however, are specific ethical issues that arise in the context of the use of AI for science and engineering. Indeed, although there has been a proliferation of ethical frameworks for AI, most of these frameworks — including the Montréal Declaration for a Responsible Development of Artificial Intelligence (2018) — do not explicitly discuss the use of AI in science and engineering contexts. One exception is the draft Recommendation on the Ethics of Artificial Intelligence, published by the United Nations Educational, Scientific, and Cultural Organization (UNESCO, 2020). This document explicitly recognizes that the use of AI in scientific and engineering practices raises "fundamental ethical concerns" and encourages awareness among scientific communities of the benefits, limits, and risks of the use of AI in these contexts, including ensuring that scientific findings generated by AI are "robust and sound" (UNESCO, 2020). In addition, the Asilomar AI principles, developed by experts at the Beneficial AI Conference in 2017, identified 23 principles for the responsible and beneficial use of AI, including research issues and ethics and values (FLI, 2017a, 2017b).

> "Ethical considerations arise at every stage in the process of using AI for scientific research, including data collection and pre-processing; the design and use of AI models trained on those data; the dissemination and publication of results; and the long-term storage, maintenance, and access to data, models, and results."

Despite the general lack of ethical guidance for the use of AI in scientific and engineering contexts, as Metcalf *et al.* (2021) point out, new forms of knowledge production may require new ethical frameworks. Indeed, insofar as existing practices in research ethics were developed before the ongoing shift towards algorithmic forms of knowledge production (such as AI), mismatches between existing ethical frameworks for research and the ethical requirements of data-driven research may be expected (Metcalf *et al.*, 2021). Ethical considerations arise at every stage in the process of using AI for scientific research, including data collection and pre-processing; the design and use of AI models trained on those data; the dissemination and publication of results; and the long-term storage, maintenance, and access to data, models, and results.

3.2.1 Ethics in Data Collection and Use

Ensuring ethical practices in the collection and use of data for the training and application of AI models is essential to the responsible use of AI in science and engineering. Two issues in particular are relevant to data collection and use: (i) ensuring that datasets are not skewed or biased such that AI models trained on them could result in discrimination against individuals or groups; and (ii) ensuring that principles such as *informed consent* are respected in data collection and use. The issue of bias and discrimination in datasets is dealt with in Chapter 5, while the issue of informed consent is examined below.

AI complicates traditional notions of human research participants and informed consent

One of the most pressing ethical issues concerning the use of AI in science arises in the context of research involving human participants. This type of research has long been subject to specific ethical requirements. In Canada, one widely used set of such requirements is found in the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS 2), authored by CIHR, NSERC, and SSHRC, most recently revised in 2018. TCPS 2 sets out the principles and policies that researchers and their institutions must adhere to as a condition of funding. Although the document does not explicitly mention AI, it does describe some ethical considerations related to big data, such as the potential for re-identification of previously de-identified data (CIHR *et al.*, 2018b).[5]

Importantly, however, the use of big data and AI tools in scientific research complicates the concept of human research participants. Whereas this term is generally understood to refer to individuals who knowingly and actively provide informed consent to participate in a research experiment, research employing big data and AI may make use of datasets containing information about individuals without their knowledge or consent.[6] This is particularly evident in relation to the use of data garnered through social media platforms (Box 3.1).

5 *De-identified* data are data from which all personal information has been removed (ICPO, 2016). However, some risk of re-identification of personal information is possible with de-identified data, which distinguishes it from *anonymized* data, which are data for which re-identification of personal information is impossible (Thompson & Lucarini, 2021). The Tri-Council Policy Statement recommends using anonymized data when possible, and de-identified data as the next best alternative (CIHR *et al.*, 2018b).

6 While TCPS 2 (Article 3.7) allows certain exemptions to the requirement to obtain informed consent before making use of personal data, such exemptions must be reviewed and approved by a research ethics board (CIHR *et al.*, 2018b). However, such review and approval does not necessarily occur in all research contexts that use big data or AI.

Box 3.1 Research Participants, Informed Consent, and Social Media Data: The Facebook "Emotional Contagion" Study

In 2014, researchers at Facebook and Cornell University published the results of an experiment that manipulated the Facebook News Feed algorithm for 689,003 people to test the impact on the emotional valence of user posts (Kramer *et al.*, 2014). However, none of the individuals whose feeds were manipulated had directly consented to participate in the study or were even aware that they were part of an experiment; rather, the intervention experiment was allowed under Facebook's Data Use Policy (Verma, 2014).

The study was criticized for not obtaining informed consent from its subjects; however, not all ethicists agree that the Facebook study was necessarily unethical. A response to criticisms of the study, signed by over 30 ethicists,[7] argued that the research neither violated the privacy of Facebook users nor was substantially different than Facebook's routine modification of its news feed algorithms (Meyer, 2014). The signatories argued that explicit informed consent was not required because the research was consistent with the reasonable expectations of Facebook users, the research imposed little or no incremental risks, and obtaining informed consent might have biased the results. However, all agreed that the study should have obtained approval from an institutional research board (Meyer, 2014). The signatories further argued that unwarranted criticism of the study "could have a chilling effect on valuable research" and could lead to similar studies being conducted in secret instead of being published.

Often, the use of personal data in the absence of explicit informed consent is defended on the grounds that such activities do not involve direct interventions in a subject's life or on the grounds that the relevant information is in the public domain (Metcalf & Crawford, 2016; Metcalf *et al.*, 2021). Such justifications rely on the assumption that what matters is *what kind* of data are collected and *how* they are collected, not *what is done* with the data after they are collected (Metcalf & Crawford, 2016). Indeed, the relative ease and low cost of collecting, storing, and re-analyzing large datasets may significantly complicate traditional concepts of informed consent because consent that is obtained at the beginning of a research

7 For transparency, the current President and CEO of the CCA, Dr. Eric M. Meslin, was among the signatories to this response. However, Dr. Meslin was not affiliated with the CCA when the response was published in 2014.

undertaking is unlikely to be sufficiently informed to adequately cover all the possible uses of an individual's personal data and the associated risks (Metcalf *et al.*, 2021). This issue is partially addressed in TCPS 2 (Article 5.5), which identifies requirements that researchers must meet in order to use information for secondary purposes without obtaining fresh consent from the participants (CIHR *et al.*, 2018b).

Many existing research ethics review boards consider publicly available data to be exempt from institutional review (Metcalf & Crawford, 2016; Leonelli *et al.*, 2021). For example, TCPS 2 (Article 2.2) provides an exemption for research that relies exclusively on information in the public domain and to which individuals have no reasonable expectation of privacy so long as there is no direct online interaction between the researchers and the subjects (CIHR *et al.*, 2018b). However, the assumption that the use of publicly available information for research purposes does not pose potential risks or harms overlooks the fact that big data and AI tools are capable of generating new and unanticipated insights or that the data may be used for purposes that consenting individuals could not have reasonably foreseen. This is particularly true when data from a variety of different sources are aggregated (Metcalf & Crawford, 2016; Metcalf *et al.*, 2021).

Publicly available data obtained from social media without the knowledge or consent of the data subjects is one example of data that could give rise to ethical challenges or harms (Box 3.2). Social media platforms provide researchers with an immense amount of data at a scale that was previously not possible, potentially offering scientifically (and commercially) valuable insights into a wide variety of topics (Leonelli *et al.*, 2021). There are several examples of scientific research making use of these data in "citizen science" initiatives (Ceccaroni *et al.*, 2019), as well as in health research (Leonelli *et al.*, 2021). However, the use of social media data for research also risks exacerbating social inequalities and injustices (Leonelli *et al.*, 2021). Moreover, there are additional concerns that the use of social media data in research could generate results that are scientifically questionable or epistemically unreliable and could even provoke public backlash or damage public perception of science (Leonelli *et al.*, 2021). There have been several attempts to develop ethical frameworks for the use of social media data. Additionally, Leonelli *et al.* (2021) identify steps that researchers can take to implement *methodological data fairness* — which focuses on "the quality and credibility of the research processes through which data are produced, gathered, pooled, analysed and interpreted" — to address epistemic and ethical issues arising from the use of data (including social media data) for research purposes.

Box 3.2 Privacy Versus Big Data: Clearview AI

Clearview AI developed an AI-based system for facial recognition that "scraped" images of faces (and associated image data) from publicly available sources, including social media, without the knowledge or consent of individuals. The company then stored that information in a database that was used by customers such as commercial organizations and law enforcement — including the RCMP — who could upload images to find matches in the database (OPC, 2021c; Thompson, 2021). A joint investigation by the Privacy Commissioner of Canada along with the Commission d'accès à l'information du Québec, the Information and Privacy Commissioner for British Columbia, and the Information and Privacy Commissioner of Alberta determined that Clearview AI violated both federal and provincial laws in its collection, use, and disclosure of personal information.[8] However, Clearview AI disagreed with the findings, failed to acknowledge wrongdoing, and has not committed to following either the recommendations or orders issued by the Canadian authorities (OPC, 2021c). The ability of Canadian authorities to penalize the company or force its compliance with Canadian laws is limited because Clearview AI is a U.S.-based company (Thompson, 2021).

Although this example falls outside the scope of this report because the information collected by Clearview AI was not used for design or discovery in science or engineering, it highlights the dangers associated with publicly available personal information and AI, and the associated difficulty with cross-jurisdictional enforcement of privacy laws related to big data.

Ultimately, the rise of big data and AI may necessitate a shift away from traditional notions of informed consent that focus on the collection of data and towards the issue of consent to the use of already existing data. This shift is partially evident in the EU's General Data Protection Regulation (GDPR), the focus of which is largely on the requirements for the processing of personal data rather than on its collection. Notably, the GDPR (Article 89-2) allows EU member states to provide exemptions from certain requirements (such as rights of access, rectification, restriction of processing, and objection) when data processing is done for scientific purposes, so long as appropriate safeguards are in place (Mourby *et al.*, 2019; GDPR, 2021).

8 In a later investigation, the Privacy Commissioner found that the RCMP violated Canadian law via its illegal use of Clearview AI's service (OPC, 2021b). The RCMP disagreed with the Privacy Commissioner's findings, although it agreed to implement the Commissioner's recommendations (Boutilier, 2021; OPC, 2021b).

In response to the increase in data-intensive research, new models of informed consent are being developed for contexts such as health data (Vayena & Blasimme, 2017). Although issues related to the use of AI for health and medical research are largely outside of the scope of this report, many of these models could be applied to scientific and engineering research more generally. For example, the *dynamic consent* model transforms consent from a static, one-time decision into an adaptive process in which participants express their preferences for which uses of their data they will and will not allow in future research via ongoing communication with researchers (Vayena & Blasimme, 2017). In addition, novel technological approaches and data governance schemes may help to protect informed consent, such as electronic consent management mechanisms and participatory data cooperatives.

The ethical use of AI for science and engineering research must focus on social harms in addition to individual harms

Although techniques such as the anonymization of personal data are intended to mitigate potential harms to individuals, the "massive aggregation of research data also turns our concept of a human subject away from individuals and toward distributed groupings or classifications" (Metcalf *et al.*, 2021). Indeed, the potential for AI to cause harm to historically marginalized groups is well known (Section 5.2). Furthermore, some have argued that notions of informed consent that focus only on individuals and not groups (and that also focus on what and how data are collected while ignoring how data are used) may create serious challenges for Indigenous communities (Section 5.3) and may contribute to a general distrust of scientific research among Indigenous Peoples (J.E. Lewis, personal communication, 2021).

> "Ultimately, the rise of big data and AI may necessitate a shift away from traditional notions of informed consent that focus on the *collection* of data and towards the issue of consent to the use of already existing data."

Despite the fact that the use of AI in scientific and engineering research can pose social and ethical risks beyond individual harms, traditional research ethics boards and institutional review boards "are designed to evaluate harms to human subjects rather than harms to human society" (Bernstein *et al.*, 2021). In response to these shortcomings, researchers at Stanford University introduced the Ethics and Society Review Board, which aims to mitigate potential negative social and ethical outcomes in AI research by functioning as a requirement for access to funding (Bernstein *et al.*, 2021). Researchers must submit a statement describing their project's potential risks to society (i.e., the society targeted by the research), social sub-groups (particularly marginalized groups), and the global community. In addition, they must identify principles for mitigating those risks and describe how those principles are implemented in the research design. These statements are then reviewed by an interdisciplinary panel that works to identify additional risks and mitigation strategies in an iterative collaboration with the researchers.

3.2.2 Ethics in the Dissemination of AI Findings

Partnership on AI (2021) has examined the question of how the results of AI research can be disseminated responsibly, with greater consideration of the downstream consequences and broader impacts of research findings produced by AI. The report provides several recommendations for three key groups: individual researchers, research leadership, and conferences and journals. For individual researchers, the proactive disclosure of additional details about the work is recommended, including its contribution and motivation, description of potential downstream consequences, and the amount of computational resources used. Research leadership is encouraged to foster early discussions of downstream consequences in internal review practices prior to undertaking the work, as well as recognizing and commending researchers who identify such consequences in their work and that of their colleagues. Finally, Partnership on AI (2021) recommends that conferences and journals expand peer review criteria to include consideration of downstream consequences and establish separate review processes to assess them. As one example of this type of practice, in addition to the disclosure checklist mentioned in Section 3.1.1, the AI conference NeurIPS also requires researchers to submit a broader impact statement (NeurIPS, 2020) that describes "both the anticipated positive and negative impacts of the paper and motivate these anticipated impacts with the proper citations" (Hecht *et al.*, 2018). However, as some peer reviewers and editors for journals and conferences may be unfamiliar with the potential downstream impacts associated with the dissemination of research utilizing AI, some degree of training to help identify potential impacts may be required. This is similar to the need for ethics education for future AI scientists and developers (Section 4.5).

3.3 The Impact of AI on Values and Social Practices in Science

The increased use of AI could affect the relative importance of different values in science, as well as its underlying social dynamics. Science and engineering are inherently social enterprises that are shaped by particular sets of values, including epistemic, social, and ethical values. Existing values may shift and new values may be introduced by the use of AI in science, while tensions may be exacerbated when they are operationalized in scientific practice. Moreover, AI is likely to affect the social practice of science in numerous ways, from selecting and conceptualizing problems, to establishing the provenance and validity of findings, to impacting the social dynamics of human-AI interaction, among others.

3.3.1 Tensions Among Different Principles and Values in AI

There are often tensions among different principles and values when it comes to the responsible and ethical use of research involving AI, such as the tension between *openness* versus *rigour* (or *prudence*) (Leonelli, 2020; Whittlestone & Ovadya, 2020). A commitment to openness requires that AI research be as open and accessible as possible; however, prudence requires that open access and public dissemination should be limited when there is a potential for misuse (Whittlestone & Ovadya, 2020). Scientific rigour requires strict monitoring of the way in which data are interpreted and used — something that is difficult to control once data are made freely available (Leonelli, 2020). Decisions about how to balance different values and resolve potential tensions when applying such principles need to consider the level of risk involved with regard to the potential for misuse versus lack of transparency, the efficacy of limiting access to prevent misuse, and future needs for preventing misuse and promoting transparency (Whittlestone & Ovadya, 2020). Other potential tensions among values include those between privacy and rights to data versus the quality of AI datasets and results (Muller, 2020); among the values underlying open science versus privacy, confidentiality, safety, and security (Science International, 2015); and between explainability and transparency versus the predictive accuracy discussed in Section 3.1.2.

Indeed, the practice of science is shaped both by epistemic values (which can include explanatory power, simplicity, or scope, among others) as well as non-epistemic social and ethical values (Douglas, 2015). Values also determine what avenues of research are pursued by scientists and what constitutes sufficient standards of evidence in a given context (Douglas, 2015). In the field of AI research, a comprehensive analysis of the 100 most highly cited papers found that the most prevalent values were performance, building on past work, generalization, efficiency, quantitative evidence, and novelty. By contrast, values related to ethical principles and user rights appeared only rarely, if at all, and almost none (2%) discussed potential harms (Birhane, 2021). Moreover, the analysis showed that, although the dominant values were *prima facie* technical in nature, the way in which they were prioritized and operationalized revealed unstated, value-laden assumptions.

3.3.2 The Sociology of Science and AI

Science is an inherently social enterprise, and the increased use of AI in this domain could change its social dynamics (OECD, 2018b). The social practices underlying the dissemination of scientific findings (publications, conferences), establishing their validity (peer review, replication), and acknowledging their provenance (crediting, citations) may all need to be revised due to the increased use of AI in scientific research (King *et al.*, 2018; OECD, 2018b). The validity of scientific knowledge is in part a socially constituted phenomenon that depends on the extent to which observations and their interpretation are agreed upon by an authoritative community of experts (King *et al.*, 2018). Similarly, the validity of machine-generated scientific knowledge will depend on the acceptance of its observations (i.e., input data) and interpretation (i.e., AI model) by the wider scientific community.

The increased use of AI in science and engineering may affect the social practice of science in many different ways

Overall, there has been little research by sociologists of science that examines the relationship between AI and human scientists or "the sociological and anthropological issues involved in human scientists and AI systems working together in the future" (OECD, 2018b). Nevertheless, the relationship between human scientists and AI can affect the practice of science at many levels, from deciding what problems or areas to investigate, to structuring or conceptualizing problems so that they are amenable to analysis by AI, to the interpretation of unusual results (King *et al.*, 2018; OECD, 2018b).

AI could also help to alleviate existing limitations and challenges in the current social practice of science — including perverse incentives around publishing and funding, human cognitive biases, "fake rigour," and insularity — by promoting collaboration rather than competition and self-promotion (The Turing Institute, 2021). However, AI could also exacerbate these challenges if it is not implemented responsibly. For example, if the use of AI became an unduly significant factor in determining the allocation of resources and funding for research, it could inadvertently close off promising areas of research that do not employ AI (Section 4.1).

AI could also shift the practice of scientific research at the epistemological level by imposing what Lowrie (2017) called *algorithmic rationality*, in which the focus of inquiry is directed at the feasibility, efficiency, and usefulness of a given algorithm or model when applied to a particular domain, problem, type of data, or dataset. Under this approach, the AI-based processes and tools by which the research is carried out become the primary focus of inquiry, thereby shaping research goals and outcomes. Furthermore, it will be important to understand how AI can contribute to scientific knowledge and how that knowledge might potentially differ from traditional, human-generated scientific knowledge (King *et al.*, 2018). This latter question is particularly relevant to the issue of whether machine-generated scientific knowledge may be or could become *epistemically opaque* — that is, unintelligible to human beings (Nickles, 2018) (Section 3.1.2).

The partnering of humans and AI in science has the potential to outperform each form of intelligence taken individually, in much the same way that teams combined of both humans and AI typically outperform both humans and AI individually in the game of chess (OECD, 2018b; The Turing Institute, 2021). Successful human-AI collaboration in science will thus require recognition and understanding of the respective capabilities and limitations of each party.

4

Implications of AI for Canada's Research System

Chapter Findings

- The use of AI in the research process is blurring the boundaries among the natural sciences, engineering, health sciences, social sciences, and humanities. Traditional, single-themed funding applications and research programs may become less relevant because domain expertise is unable to properly review the interdisciplinary nature of AI-driven science.
- AI's ability to predict the impact of scientific research may be of use to governments and funding agencies when making decisions about resource allocation and funding opportunities for the scientific community.
- AI can support the peer review process used to evaluate prospective scientific applications; however, this will need to be carefully tested to ensure that it does not result in unintended negative consequences.
- The use of open data in research is gaining momentum, but incentives to share are not universal. New policies and investments aim to overcome barriers and establish a harmonized data landscape.
- AI reconceptualizes research integrity and the governance of research conduct within the scientific community.
- Educating future researchers in the age of AI will not be solely a matter of broadening technical knowledge and skills. Academics are recognizing the need to teach future scientists how to think through ethical dilemmas associated with the development and use of AI.

In Canada, the allocation of public resources for research performed outside of government is primarily overseen by three agencies: CIHR, NSERC, and SSHRC (i.e., the Tri-Agencies), which divide responsibility into the general areas of health, natural sciences and engineering, and social sciences and humanities, respectively. The academic research funding system and its oversight comprise several stages of decision-making involving researchers, institutions, and granting agencies. Granting agencies make funding available by calling for proposals from the scientific community. Researchers then decide whether to compete by preparing and submitting a proposal. Granting agencies typically award funding based on merit as determined by peer review. Lastly, post-grant management involves decisions by the granting agency, the successful researchers, and their institutions. As a decision-making technology, the application of AI in any of these stages will likely have a significant impact on the research funding system and its governance.

4.1 The Use of AI in Allocating Research Resources

The use of AI in the research process is blurring the boundaries among the natural sciences, engineering, health sciences, social sciences, and humanities, challenging the utility of traditional single-discipline funding competitions and research programs. Thus, AI is amplifying the Advisory Panel on Federal Support for Fundamental Science's observation that stronger supports are necessary for multidisciplinary research (Naylor *et al.*, 2017). The humanities and social sciences will have an important role to play in scientific and engineering R&D that uses AI. Existing traditional disciplinary divisions in research funding may need to be rethought to ensure fair and appropriate assessment of research using AI.

Research funding is a complex process of balancing political, economic, and scientific interests to determine whether to provide funding and in what fields. For instance, prioritizing knowledge translation and applied research — which are seen to have more immediate potential commercial returns — over basic (i.e., curiosity-driven) research has resulted in a restructuring of Canada's research ecosystem (Naylor *et al.*, 2017). Increasingly, funding agencies are expected to provide funding opportunities for a growing population of researchers with comparatively less money (CCA, 2021a). Efforts to respond to these circumstances by revamping the structures supporting funding decisions may give rise to other issues, such as underfunding certain researchers due to their career stage or the novelty of their field (Reardon, 2015), or lower-quality reviews prompted by virtualizing the peer review process (Webster, 2015; Woodgett, 2018). Following the adoption of AI systems by publishers and conference organizers (Heaven, 2018; Hutson, 2021), it is reasonable to suggest that these systems will be a part of such resource allocation efforts in the future.

By predicting research impact, AI systems may increasingly determine the trajectory of scientific discovery

AI's ability to predict the impact of scientific research (Section 4.3) may be of use to governments and funding agencies making decisions about how and whether to allocate resources to the scientific community. AI may also inform and identify research priorities for these funders. As Weis and Jacobsen (2021) suggest, an AI program (e.g., the Dynamic Early-warning by Learning to Predict High Impact program) could be used to "aid in designing funding strategies" by filling holes in the research terrain with "new research program opportunities designed to optimize for connections of predicted high impact." This use of AI gives rise to concerns that it potentially creates a self-fulfilling prophecy: if more grant money is allocated to a certain field because it has been predicted to be highly impactful, more researchers will work in that field and presumably produce more outputs that will be recorded as "impact" (Chawla, 2021). This does not necessarily mean that the allocation of funding best serves the research system or Canadian discovery as a whole.

As AI's role in the discovery cycle becomes more central, the application process for researchers is expected to change

Research grant awards support cycles of scientific discovery — the deployment of background knowledge to formulate hypotheses, the testing of hypotheses by way of experimentation, the interpretation and analysis of experimental results, and the eventual contribution of final theories to the background knowledge. The integrity of this cycle may be compromised at each stage by human-specific limitations and flaws, such as incompleteness, error, and bias. The growing competitiveness of the research funding system can have a counterintuitive effect of inducing such conduct (The Turing Institute, 2021). As human limitations are increasingly apparent in managing high volumes of data requiring rapid and accurate processing, the need for solutions grows, such that gaps where these limitations occur might conceivably be filled by an "AI fully automated system" (The Turing Institute, 2021). These changes will likely transform the funding application process itself.

AI has the potential to drive scientific investigation by allowing for automated hypothesis generation, experiment design, experimentation, interpretation, and analysis

Research grant proposals are typically investigator-driven endeavours, as opposed to strategic or sponsor-driven; that is, the hypothesis proposed to be studied originates from the investigator's understanding of the field's background knowledge. Yet the continuous expansion of background knowledge has made simply keeping abreast of one's field of expertise challenging, a circumstance that will necessitate "automated paper review" (The Turing Institute, 2021). AI is being developed to provide reviewers with a paper's general summary, contribution summary, writing quality analysis, and related works to reduce the time burden of peer review (Roberts & Fisher, 2020). The use of AI in science also makes possible automated hypothesis generation, experiment design, experimentation, interpretation, and analysis — all currently in use to some degree. For example, AI is being used to read, interpret, and infer from the human genome to predict which compounds are promising candidates to treat mutations causing disease (Wainberg *et al.*, 2018). Deep Genomics, a Canadian company harnessing AI to discover new drug candidates, suggests that the "future of drug development will rely on AI, because biology is too complex for humans to understand" (Deep Genomics, 2021). In this way, AI's application to scientific discovery is more than just another technology to eliminate the drudgery of repetitive tasks, as was the case with calculators and computers. Instead, AI will be more or less driving the investigation.

Competition for funding could become less a matter of merit and more a matter of access to AI resources. Consider, for example, the impact of a robotic arm that can perform high-risk toxicology tests in one day that would take a human one year to complete. Access to such a technology has been described as transformative to our understanding of toxicology, given the rapidity of testing (Bogue, 2012). Presumably, funding applications from toxicologists who have a robotic arm in their lab will come across as stronger, more complete proposals than others that do not. AI may take this kind of advantage to another level by going beyond lab automation to make possible "robotic scientific discovery" (Bogue, 2012), or, as Sparkes *et al.* (2010) put it, the "robot scientist." If biology becomes too complex, researchers with access to AI models that can make novel discoveries and more accurate predictions will likely give biologists using AI an advantage in funding competitions.

> "AI's application to scientific discovery is more than just another technology to eliminate the drudgery of repetitive tasks, as was the case with calculators and computers. Instead, AI will be more or less driving the investigation."

A consequence of this scenario would be the concentration of funding to the economically advantaged researchers and organizations that have access to AI. Of course, well before the use of AI in discovery work, the research funding system was guided by factors other than the quality of research proposals (Laudel, 2006). As Robert Merton argued over 50 years ago, the social processes of resource allocation in science give rise to a Matthew effect, whereby previous funding success leads to more funding in the future (Merton, 1968). The use of AI in discovery risks introducing a new sort of Matthew effect, one that is not only resource-mediated but also intelligence-mediated. In this context, the idea of "merit" — a fundamental principle governing the distribution of competitive research funds — would be a quality increasingly defined by non-human ability.

4.2 Adapting Peer Review Systems to AI

As competition for funding grows, the peer review process is put under increased strain, pushing funding agencies to make changes to their peer review processes to decrease the time, resources, and effort deployed. Publishers and funding agencies are using automation to screen submissions for plagiarism, for formatting compliance, and to assign reviewers (Charlin & Zemel, 2013; Grant, 2017; Checco *et al.*, 2021). As such, AI has been deployed as a tool for maintaining scientific quality in the early stages of the review process.

AI can support the peer review process, but issues related to conflicts of interest and complexity will require careful attention

Reviewer assignment is a detail-intensive, repetitive task amenable to AI solutions. Publishers are beginning to screen submissions for conflicts of interest among investigators, editors, and reviewers (Chawla, 2020). However, although AI may be superior at identifying connections among scientists, there are nuances in conflicts of interest that AI may not necessarily be capable of detecting. These conflicts may be subjective, depending on the size of the field and the nature of previous contact among scientists (Chawla, 2020). For example, although an applicant should not be evaluated by a reviewer from the same institution, this may be permitted if the circumstances are appropriately justified. Conflicts are not identified in a binary way as either present or absent; they are identified by assessing their severity (Lo & Field, 2009). Judgments must be made about "the likelihood that professional decisions made under the relevant circumstances would be unduly influenced by a secondary interest" and "the seriousness of the harm or wrong that could result from such influence" (Lo & Field, 2009).

Peer review decisions informed by automated systems may also be obscured by the algorithm's own opacity (Checco *et al.*, 2021). Thus, instead of resolving the black box problem of peer review (Oransky & Marcus, 2017), AI may simply replace one black box with another. Whether informed by AI or not, funding agencies need to ensure that their decisions are interpretable by applicants and the public if funding outcomes are to be transparent and ultimately trusted. For some applications, transparent machine learning models that are interpretable by humans can be constructed that arguably perform just as accurately as black box models (Rudin, 2019; Rudin & Radin, 2019).

AI will have to be programmed to support an equitable, diverse, and inclusive peer review system

Equity, diversity, and inclusion (EDI) in the research system are priorities established by the Tri-Agencies. One intended outcome of the Tri-Agency EDI Action Plan for 2018–2025 is the inclusion of diverse participants in review committees (CIHR *et al.*, 2018a). To achieve this, policies and processes are to be put in place to ensure that peer review is inclusive and reflective of Canada's diversity (CIHR *et al.*, 2018a). AI employed to inform reviewer selections will need to be capable of making nuanced selection decisions based on considerations of expertise and diversity. There are concerns that AI could exacerbate the marginalization of traditionally under-represented groups in the research system (Chapter 5). The integration of AI systems and technology into the peer review system could mitigate this issue, but it cannot be presumed. It will require careful implementation, taking into consideration the limitations of AI.

> "Replacing humans with an AI-based system to make peer review decisions is unlikely to remove inherent or implied biases."

The application of AI in peer review may do more than streamline administrative tasks that take up time and are prone to human error. It may also be used to determine submission readability, relevance, and adherence to formatting requirements, thereby preventing reviewers from making decisions based on first-impression biases rather than scientific merit (Checco *et al.*, 2021). However, AI may have its own biases. As Checco *et al.* (2021) point out, AI is trained with data from the past and is therefore inherently conservative. There is a risk that problems signalled by automated submission screening may affect the role of the reviewer, perpetuating the values and expectations that are implicit in the data on which the system is trained (Checco *et al.*, 2021). This may disadvantage groups that have been under-represented in the literature.

Replacing humans with an AI-based system to make peer review decisions is unlikely to remove inherent or implied biases. Given the reliance of many AI systems on machine learning techniques that extract patterns from the data used, dataset integrity is critical to the quality of the AI system (Muller, 2020). In other words, an AI system working with a biased dataset will turn out to be a biased AI system. Yet technical bias is only one form of potential bias in AI systems used for peer review. AI systems "are shaped by the environments in which they are built and the people that build them" (West *et al.*, 2019). A "diversity crisis" in the social spaces where AI systems are built may cause differential treatment of more people as these systems play increasingly important roles in more domains (West *et al.*, 2019). As EDI is prioritized in peer review, the AI systems deployed for this purpose will need to be based on equitable, diverse, and inclusive datasets and design principles.

4.3 Measuring Research Impact

AI may be capable of predicting the impact of scientific research better than humans or currently employed indicators such as citation count (i.e., by anticipating the number of future studies that will refer to the funded research project). In the same way that the predictive capacity of AI might indicate promising fields of study (Wang & Barabási, 2021; Weis & Jacobson, 2021), funding agencies may use this technology to inform decisions about grant renewal. Of course, the actual impact of research funding is only realized years after the decision to award funding is made. Attempts to pre-determine impact using factors other than the scientific work itself are at risk of coming to faulty conclusions. For example, using journal impact factors to evaluate research assumes that the quality of the journal is representative of the quality of the article — a flawed assumption, given the different ways that a journal's impact factor can be manipulated (Seglen, 1997).

By "considering a broad range of features and using only those that hold real signal about future impact," Weis and Jacobson (2021) claim that more sophisticated AI reduces the latent systemic biases in the system. However, there is nonetheless a risk that AI used to predict research impact is simply employing "ever more sophisticated versions of basically useless indicators" (Seglen, 1997). Regardless of indicator utility, the predictive capacity of AI is typically based on historical data. Thus, bias against those without research profiles (e.g., early-career investigators) is perhaps unavoidable when relying on AI systems to inform funding decisions based on research impact. As argued by Sydney Brenner (1995), unless the technology can know and read the scientific content of a paper, AI cannot take the place of an expert review (Seglen, 1997). As recognized in the San Francisco Declaration on Research Assessment, research needs to be assessed on its own merits (DORA, n.d.).

New policies and investments incentivizing data sharing can help realize the potential of AI in the discovery process

AI applications for science and engineering will frequently require access to data that will not necessarily fall within the scope of open data initiatives launched by government agencies. Instead, they will rely on the availability of data resulting from academic research. The development and growth of AI for science and engineering depends on the ability of researchers to access, manipulate, and store research data, resulting in requirements for data storage and data-sharing infrastructure, and the definition or refinement of data standards and governance frameworks (The Royal Society, 2019).

Several core challenges exist for promoting the uptake of open-data practices outside of the public sector. Funding agencies and high-profile scientific journals have encouraged or mandated open data to varying degrees (cOAlition S, 2021; Yeston, 2021), but professional incentives for researchers to disclose data are otherwise unclear when measured against the risks of their data being misused and their authorship uncredited (TMS, 2017). Although the level of compliance with open-data principles by researchers in Canada is difficult to pinpoint, it is inconsistent across disciplines and lower than in several other jurisdictions (Larivière & Sugimoto, 2018), despite policies mandating open access implemented by the major public funding agencies (CIHR, 2019). Moreover, the Tri-Agency open-access policies focus on publications rather than data — for example, NSERC does not require open data to comply with its policy (NSERC, 2014) — such that the availability of data may hinge on the policies of the journals where researchers publish. A 2021 update to the Tri-Agency research data management (RDM) policy states that the requirement of open data is not mandated, though researchers are nevertheless asked to identify reasons (e.g., ethical, legal, or commercial) that would preclude data sharing in RDM plans they submit as part of funding applications (GC, 2021d). Grant recipients will soon be required to store code, data, and metadata underpinning publications in a digital repository, but researchers will have some leeway in defining which research data are considered relevant. All institutions administering research grant funds will be required to develop RDM strategies (GC, 2021d).

> "The development and growth of AI for science and engineering depends on the ability of researchers to access, manipulate, and store research data, resulting in requirements for data storage and data-sharing infrastructure, and the definition or refinement of data standards and governance frameworks."

Norms and practices regarding the production and use of data — such as volume, formats, and workflows — differ across disciplines and institutions. The large number of stakeholders with diverse priorities and needs complicates the modernization of RDM in Canada (Baker *et al.*, 2019). Repositories act as key elements of infrastructure for this purpose but are tasked with addressing several challenges, including harmonizing standards, offering storage capacity, and creating communities of practice. Existing repositories tend to operate independently and are organized according to the needs of the communities that use them the most. Incentives to introduce broader interoperability may be lacking (Baker *et al.*, 2019). Data repositories are funded inconsistently[9] and frequently through competitive processes, tying their existence to projects that require compliance with the policies of the granting agency as a condition of funding. This can undermine resiliency because repositories might be created and maintained by individual researchers and therefore be susceptible to funding gaps (Baker *et al.*, 2019).

Several recent initiatives pertaining to research data are underway, with implications for improving access to data across the Canadian research ecosystem. In 2019, the Digital Research Alliance of Canada ("the Alliance," formerly the New Digital Research Infrastructure Organization) was created to lead the establishment of an integrated digital environment for research in Canada, with a focus on computing, RDM, research software, and cybersecurity (NDRIO, 2021a; The Alliance, 2021). The Alliance held a consultation with researchers to confirm the diversity of needs and priorities across disciplines, but the overall top-identified need expressed was for data repositories (NDRIO, 2021b). The Alliance and Compute Canada are collaborating to establish the Federated Research Data Repository — a single platform for data repositories shared across multiple institutions — but this initiative was only recently funded (FRDR, 2021).

Baker *et al.* (2019) suggest that Canada lags behind its international peers in terms of both human and financial resources dedicated to RDM and that an opportunity was missed to participate in international discussions regarding data standards,[10] thereby fostering a digital research infrastructure ecosystem that supports Canadian R&D activities in AI. Large, open repositories containing data are instrumental in advancing AI for design and discovery across numerous disciplines (U.S. NRC, 2014; De Luna *et al.*, 2017; TMS, 2017). Some international jurisdictions have already established repositories of this scale, distributed over multiple pieces of regional infrastructure; for example, the Australian Research Data Commons was formed in 2018 to consolidate previously distinct organizations for data infrastructure as part of Australia's 2016 National Research Infrastructure Roadmap (ARDC, 2021). More recently, the European Open Science Cloud (EOSC) launched a hub for FAIR-compliant data and publications associated with research activities in the EU (EOSC, 2021). Looking ahead, the EOSC directly names machines among the main users of its infrastructure in its published strategy documents (EOSC, 2021).

9 For example, long-term data storage is ineligible for support through the Canada Foundation for Innovation, and the Canadian research environment is lacking in storage of this type (Baker *et al.*, 2019).

10 Domestic standardization efforts are also underway (SCC, 2020), with additional funding in the 2021 federal budget (GC, 2021c).

4.4 Research Integrity and the Governance of Research Conduct

AI reconceptualizes responsibility within the scientific community, challenging existing research conduct policies

Research funding is governed in part by ethical codes with which researchers and their institutions are required to comply. The Tri-Agency Framework for Responsible Conduct of Research imposes responsibilities on researchers funded by the agencies to support research integrity and address misconduct. Researchers are responsible for providing accurate and complete information to the agencies when applying for funding (CIHR *et al.*, 2016). The Government of Canada has a Model Policy on Scientific Integrity that guides departments and agencies in ensuring that their intramural research complies with various ethical principles and rules (GC, 2021e). Also, as discussed in Chapter 3, rules specifically governing research involving human subjects regulate researcher conduct. Given the promise of AI to increase accuracy and consistency, the requirements of researcher responsibility are subject to change. If AI systems become the standard for completing scientific tasks, expectations surrounding the responsibility for accuracy and completeness may eliminate human-completed work from consideration. This could remove applicants without access to AI from consideration for funding based on presumed irresponsibility.

> “Who, or what, exactly is responsible for breaches of policy against fabrication, falsification, plagiarism, and so on, when such conduct may be traced back to tasks completed by AI?”

The introduction of AI into the research process complicates the idea of a breach of the RCR framework’s policy. Issues discussed in Chapter 3, such as reproducibility, explainability, and accuracy, have practical significance here. Who, or what, exactly is responsible for breaches of policy against fabrication, falsification, plagiarism, and so on, when such conduct may be traced back to tasks completed by AI? Although AI is serving to enforce scientific integrity by detecting these breaches of responsible research, it could also be used to perpetrate irresponsible conduct by researchers. It might also be possible for an AI algorithm itself to cause such breaches. A model of “responsible AI” that provides mechanisms to “enable AI systems to act according to ethics and human values” (Dignum, 2020b) will be important to sustaining research integrity.

AI systems used in the research process will need to be designed such that they conduct their research tasks in ways that align with the research ethics governing funding competitions. Designing responsible AI will require explicit and systematic considerations about principles of accountability, responsibility, and transparency (Dignum, 2020b). This necessitates researchers and developers who are "trained to be aware of their own responsibility where it concerns the development of AI systems with direct impacts in society" (Dignum, 2020b).

For instance, research involving Indigenous communities (Section 5.3) gives rise to responsibilities for researchers and their AI models that resonate with broader concerns about relations between AI and the human subjects of research (Section 3.2). Community engagement is particularly important to any research affecting Indigenous communities (CIHR *et al.*, 2018b). Given the long history of science and technology being used in ways that have not benefited Indigenous Peoples or communities (CIHR *et al.*, 2018b), the use of AI in research involving Indigenous Peoples may be met with suspicion. To be successful, AI will need to align with *Indigenous Protocol* — the customs, lore, and codes governing behaviour — and find a place within existing circles of relationships (Lewis, 2020).[11] The use of AI in research involving Indigenous Peoples will therefore need to be understood by those involved, such that the cultural presuppositions encoded in the technology are recognized and capable of reorganization (Lewis, 2020).

4.5 Training and Skills Development

The discovery process of the future will require developing AI expertise

As the application of AI expands, more fields will be moving goalposts with respect to their expectations for researchers' skills and knowledge (Fleming, 2018). Educational programs in the sciences will change in the future to reflect the shifts in necessary skills prompted by AI. Skills development and training will likely be needed to adapt the workforce to the changes generated by the use of AI in the lab (Lane & Saint-Martin, 2021). For example, researchers may require training to learn how to effectively use data science tools (Ezer & Whitaker, 2019). According to an analysis by the OECD (2021b) using data averaged from 2015–2020, Canada ranks fifth in terms of national AI skills penetration, above the G20 average (ninth), but behind India, the United States, China, and Germany.

11 On the importance of accepting Indigenous traditions for developing a more complete understanding of the implications of Western thought, see Borrows (2010, 2012).

Decisions will need to be made about whether to broaden degree programs and at what level of study. Although some believe that PhDs will need to be more interdisciplinary (for example, by requiring biologists to understand not only biology but also machine learning and other computer science concepts), others believe undergraduate programs will be made broader while PhDs remain more focused on deep, disciplinary skills (Fleming, 2018). Universities are not solely responsible for educating future AI developers. Leaders in AI are leveraging industry to support practical training initiatives in a way that benefits students, businesses, and ultimately Canada's AI ecosystem (Box 4.1).

> "Educating future researchers in the age of AI will not be solely a matter of broadening technical knowledge and skills. Academics are recognizing the need to teach future scientists how to think through ethical, cultural, and social dilemmas associated with the development and use of AI."

Educating future researchers in the age of AI will not be solely a matter of broadening technical knowledge and skills. Academics are recognizing the need to teach future scientists how to think through ethical, cultural, and social dilemmas associated with the development and use of AI (Dignum, 2020a). For example, the University of Toronto will be initiating a pilot program that embeds ethics modules into its undergraduate computer science program, reconfiguring courses to spend time teaching students how to identify the ethical questions and broader societal implications of future technologies (UofT, 2021b). Following the strategy of Harvard University's Embedded EthiCS program, the objective is not necessarily to improve abstract thinking by computer scientists and engineers but rather to teach students to include ethical considerations in the design of new technologies. As Bezuidenhout and Ratti (2020) suggest, curricular shifts to emphasize questions that data scientists will face in their professional activities may affect character development by internalizing ethics training that can be applied to microethical contexts.

Box 4.1 Mitacs and Scale AI

Mitacs is a national non-profit organization in Canada that provides funded internship and fellowship opportunities by connecting students with businesses and the public sector, thereby creating post-secondary training opportunities to develop recent graduates' practical expertise. Mitacs has committed to supporting the AI ecosystem by connecting small- and medium-sized enterprises (SMEs) that want to launch AI-related projects with graduate or postdoctoral students who can provide expertise and advice about how to get started.

Canada's AI supercluster, Scale AI, is also investing in AI education in Canada. By subsidizing accredited training programs for full-time workers, Scale AI aims to enhance the current workforce's skills in digital-enabled activities. Programs dedicated to generating interest in science, technology, engineering, and mathematics (STEM) programs among youth also provide investments in the future workforce.

The current organizational structure of research funding, which divides health, science and engineering, and social sciences and humanities research per the Tri-Agency's mandate, will need to adapt to "encourage, facilitate, evaluate, and support multidisciplinary research" (Naylor *et al.*, 2017). An example of this adjustment is the Tri-Agency's New Frontiers in Research Fund, launched to support interdisciplinary research in Canada (SSHRC, 2021).

5

Societal Implications of Research that uses AI

Chapter Findings

- A lack of trust in AI used in other domains could create challenges for its uptake in science and engineering. Alternatively, uncritical trust in AI could lead to an overreliance on the research results generated by AI.
- Given the well-known problem of AI systems perpetuating bias and discrimination against historically marginalized groups, it will be essential that its use in science and engineering research does not generate biased results that could create discriminatory outcomes.
- The significant inequalities in access to AI resources, infrastructure, and employment opportunities that currently exist could negatively impact EDI progress in scientific and engineering research that uses AI.
- Although the nature and extent of the impact of AI on the science and engineering labour market are unclear, many jobs will be transformed by its increased prevalence.
- AI systems can have a significant environmental impact due to the energy required to power computational infrastructure. However, mitigation of these impacts is an ongoing area of study, and AI holds promise for scientific and engineering research that could help to reduce environmental damage and climate change.
- Despite the recognition that robust security measures are needed to protect AI systems from various kinds of attacks, there is little research that specifically examines the role of cybersecurity in the context of AI for science and engineering.

The increased use of AI in science and engineering research will undoubtedly affect society, the economy, and the environment. Social trust in AI technology could impact its uptake in science and engineering, and its responsible use will require managing the risks of perpetuating bias or discrimination with respect to both the findings generated by AI systems and access to the technology itself. The increased use of AI in science and engineering could also impact labour markets and skills demands in those fields, as well as potentially increase greenhouse gas (GHG) emissions. In addition, cybersecurity measures will be required to ensure the safe and responsible use of AI in the context of science and engineering.

Many or most of these issues — including social trust, bias and discrimination, equitable access to resources, labour market disruption, environmental degradation, and security — are social challenges that pre-date both the general introduction of AI into various areas of society and the use of AI in the specific context of science and engineering research. The introduction of AI should therefore be understood as amplifying these pre-existing issues while also creating new related complications and challenges to them.

5.1 Trust in AI: Automation Bias and Algorithmic Aversion

A lack of trust in AI may act as a barrier to its adoption in science and engineering contexts. For example, O'Connor *et al.* (2019) suggest that a lack of trust in AI is a current barrier to its use for systematic reviews in research. To overcome this barrier, AI systems will need to build a trusted evidence base by transparently demonstrating successful, reproducible results. Trust is also required for effective human-machine collaboration, without which team members may have to "spend unnecessary energy and time to re-inspect work and revalidate decisions" (Hou *et al.*, 2021).

Scientific discoveries or engineering designs generated by AI may be unfairly perceived as either superior or inferior to their human-generated equivalents

Differences in the perceived accuracy or utility of AI compared to humans in the context of science and engineering may demonstrate *automation bias* (Lopez *et al.*, 2019), where the results generated by machines are perceived as more trustworthy than those generated by humans; or *algorithmic aversion* (Dietvorst *et al.*, 2015), where the results produced by AI are perceived to be less trustworthy than human-generated ones.

For example, a study by Lopez *et al.* (2019) found that people were somewhat more likely to perceive AI-designed sketches of boats as more functional than human-designed ones. This effect was compounded when sketches were explicitly labelled as having been generated by a computer or by a human, suggesting that participants' perceived functionality of the sketches may demonstrate automation bias. Similarly, research by Logg *et al.* (2018) found that individuals consistently valued the same advice more when it was labelled as generated by an algorithm compared to a human being. In contrast, research by Dietvorst *et al.* (2015) demonstrated that, compared with human-generated results, people are far more likely to lose confidence in algorithmically generated results after seeing the system err, even if they also see that the algorithms reliably outperform humans. Indeed, participants' confidence in an algorithmic model consistently decreased after seeing it make small mistakes, while seeing a human make large mistakes did not consistently decrease their confidence in the human (Dietvorst *et al.*, 2015).

Notably, in both the Lopez *et al.* (2019) and Dietvorst *et al.* (2015) studies, the AI consistently outperformed its human competitors, meaning that, in both cases, trusting AI was the best strategy to ensure accuracy. However, accuracy alone may be insufficient to build trust in AI; the transparency and explainability of AI results (Section 3.1.2) may be critical to building trust (Hou *et al.*, 2021). Furthermore, Logg *et al.* (2018) found that people were more willing to rely on advice generated by an algorithm over human advice even in cases where the system was a black box and suggest that, for laypersons, additional information could do more to increase doubts than to increase trust. Nevertheless, frameworks for trust in AI developed by the European Commission (2019) and the OECD (2021a) (Table 5.1) consider transparency and explainability to be vital for trust in AI, particularly for automated decision-making. Interpretability will likely be necessary if people are to trust AI in decision-making contexts related to science and engineering, such as funding decisions and peer review (Section 4.1).

AI may be perceived as (un)trustworthy for different reasons in different contexts

Different contexts give rise to different trust-related challenges for AI (AI HLEG, 2019). For example, AI systems that make music recommendations do not produce the same concerns about trust compared to AI systems that recommend critical medical treatments or feature in government decision-making. Regardless, even seemingly innocuous AI systems can raise concerns about trust related to their practices for data collection and use, particularly when those data are used by other AI systems. Thus, the question of trust in AI must be understood in the unique context of its use in science and engineering research and discovery. However, there is little to no research on the issue of trust in AI in these contexts. Nevertheless, the lack of public trust in AI in other domains — such as the justice system, healthcare, labour market, and social media — might negatively affect perceptions of the trustworthiness of AI in science and engineering. If this is the case, then building trust in AI for science and engineering may require building trust in AI in these more public-facing domains.

Table 5.1 Principles for Trustworthy AI

European Commission	OECD
• accountability • diversity, non-discrimination, and fairness • environmental and societal well-being • human agency and oversight • privacy and data governance • technical robustness and safety • transparency	• accountability • human-centred values and fairness • inclusive growth, sustainable development, and well-being • robustness, security, and safety • transparency and explainability

(AI HLEG, 2019; OECD, 2021a)

Both the OECD and the European Commission have developed frameworks for trustworthy AI. Both sources identify technical and non-technical methods and tools that can be used to implement these principles.

5.2 Managing Bias and Discrimination in AI Systems

The responsible use of AI in science and engineering must avoid perpetuating bias and discrimination against individuals or groups. As noted in Chapter 3, discriminatory outcomes are a well-known challenge with AI, and its use in various contexts has been found to discriminate against historically marginalized groups. For example, AI systems have been found to discriminate against women (Dastin, 2018) and people with mental disabilities (Fruchterman & Mellea, 2018) in hiring decisions. The use of AI has also resulted in discrimination against Black people in healthcare settings (Obermeyer *et al.*, 2019) and in risk assessments for recidivism in the criminal justice system (Dressel & Farid, 2018). AI was found to discriminate against students from historically lower-performing schools when assigning grades (Satariano, 2020), and against Black people (Sap *et al.*, 2019) and people with disabilities (Hutchinson *et al.*, 2020) when attempting to detect speech deemed toxic or hateful. These types of discriminatory biases in AI systems could present novel challenges for anti-discrimination laws in Canada, such as in situations concerning access to confidential data in the context of litigation surrounding bias and discrimination caused by AI. See Chapter 6 for a discussion of related legal issues.

The risks of discriminatory AI in the context of science and engineering depend on the purpose for which it is used

Generally, the use of AI for science and engineering will not directly affect individuals or groups, at least insofar as its use is for the purpose of *discovery* and not *decision-making*. However, as noted in Chapter 4, AI may indeed be used for decision-making in scientific and engineering contexts, such as peer review and funding decisions. Even in the context of discovery, it will be important to ensure that the design and use of AI systems do not negatively affect individuals indirectly, for example, by producing inaccurate results that are used by others to justify discriminatory decision-making. Of course, this issue also arises with respect to results in science and engineering that are not generated by AI. However, the opacity of some AI systems means that the process that produced the results in question may not be transparent to the users of the findings who make decisions, nor even to the researchers who designed the system (Section 3.1.2). Thus, the responsible and ethical dissemination of the results of machine-generated research findings (Section 3.2.2) must be considered to avoid perpetuating discrimination.

Biases in training data can perpetuate discrimination, but they may be ameliorated through a greater focus on data provenance, accountability, and transparency

Discriminatory outcomes often reflect biases in an AI model's training data, which may reflect historical biases. In other cases, discriminatory outcomes may result from excluding classes of individuals who do not generate a lot of data, such as people living in the rural areas of low-income countries. Additionally, the data might simply be low-quality or error-ridden (WEF, 2018). However, historical and sampling bias are not the only ways that datasets can be biased; for example, data are often processed in a variety of ways before they are used in AI applications, and the subjective decisions involved in such preprocessing can have significant ramifications for the behaviour of AI systems (Veale & Binns, 2017).

> "Choices are not neutral but rather inescapably value-laden and reflect the worldviews of the individuals developing the model."

One way that experts have suggested managing biases and avoiding discriminatory outcomes in AI research is through *data provenance* — that is to say, documenting the history and process of a dataset's selection, construction, and use (WEF, 2018; West *et al.*, 2019). To operationalize this process, Gebru *et al.* (2021) have proposed *datasheets for datasets*, which would document the various features of datasets, such as the motivation for their creation, their composition, their collection process, the preprocessing and labelling process, their recommended uses (including distribution and maintenance), and any other relevant features. This would help both creators and users of datasets by encouraging creators to engage in "careful reflection on the process of creating, distributing, and maintaining a dataset, including any underlying assumptions, potential risks or harms, and implications of use" (Gebru *et al.*, 2021), while also providing users with the information necessary to make an informed decision about the use of a dataset. Datasheets for datasets may also be useful to "policy makers, consumer advocates, individuals whose data is included in those datasets, and those who may be impacted by models trained or evaluated on those datasets" (Gebru *et al.*, 2021). See Section 3.1 for a discussion of how datasheets for datasets could help to facilitate reproducibility in AI research.

> "Bias and discrimination in AI systems are often the result of systemic issues of power and inequality in the institutions that produce those systems, a challenge that is exacerbated by the overall lack of gender and racial diversity in the field of AI research."

However, data provenance alone will not solve the problems with discrimination in AI. Even in cases where AI models are trained on high-quality, non-biased datasets, they can still produce discriminatory outcomes if the model is inappropriate for the study, if its parameters are inadvertently discriminatory, or if there is a lack of human oversight, a lack of transparency, or even intentional discrimination (WEF, 2018). Thus, in addition to data provenance, documentation of the subjective choices made by researchers in the process of developing an AI model can help to identify potentially discriminatory biases (Selbst & Barocas, 2018). Indeed, these choices are not neutral but rather inescapably value-laden and reflect the worldviews of the individuals developing the model (Veale & Binns, 2017). Thus, accountability is crucial to identifying and mitigating discriminatory outcomes (Berendt & Preibusch, 2017).

Researchers working in areas such as discrimination-aware data mining and fairness, accountability, and transparency in machine learning have proposed different computational techniques and technical approaches to identify and mitigate discriminatory biases in datasets and AI systems (Veale & Binns, 2017).[12] However, technical measures alone are not sufficient to avoid discrimination in AI; rather, human choices and actions around AI development and use can result in discriminatory outcomes even after datasets have been assessed for biases (Berendt & Preibusch, 2017).

12 See Stoyanovich *et al.* (2016) and Hajian *et al.* (2016) for reviews of the technical details of some of these techniques.

Finally, AI can also have discriminatory outcomes due to a lack of diversity in the development of an AI system; as West *et al.* (2019) put it, "tackling the challenges of bias within technical systems requires addressing workforce diversity, and vice versa." Bias and discrimination in AI systems are often the result of systemic issues of power and inequality in the institutions that produce those systems (West *et al.*, 2019), a challenge that is exacerbated by the overall lack of gender and racial diversity in the field of AI research (Section 5.4).

5.3 Impact of AI on Indigenous Communities

The Indigenous data sovereignty movement aims to advance Indigenous self-determination and governance with respect to data and to address biases in data that may discriminate against Indigenous Peoples

In Canada and elsewhere, Indigenous communities are disproportionately impacted by biases in datasets. For example, the lack of disaggregated health data in Canada obscures differences in healthcare outcomes for Indigenous Peoples (BCOHRC, 2020), while socio-economic data for Indigenous populations living in urban areas or small communities are often skewed due to relatively small sample sizes (Steffler, 2016). This has led to Indigenous-led movements for Indigenous data sovereignty, which refers to the right to self-determination and governance over access, ownership, and use of data and data analyses (Rainie *et al.*, 2017; Crawford *et al.*, 2019). The concept of Indigenous data sovereignty has gained increased attention as "big data, open data, open science, and data reuse become an integral part of research and institutional practices" (Carroll *et al.*, 2020). Examples of organizations and initiatives promoting Indigenous data sovereignty include:

> "In Canada and elsewhere, Indigenous communities are disproportionately impacted by biases in datasets."

- the U.S. Indigenous Data Sovereignty Network, established in 2016, which aims to "link American Indian, Alaska Native, and Native Hawaiian data users, tribal leaders, information and communication technology providers, researchers, policy-makers and planners, businesses, service providers, and community advocates together to share stories about data initiatives, successes, and challenges, and resources" (USIDSN, n.d.);
- the Local Contexts initiative, established in 2010, which aims to help manage Indigenous data issues in digital environments, focusing on "increasing Indigenous involvement in data governance through the integration of Indigenous values into data systems" (Local Contexts, n.d.);

- the Global Indigenous Data Alliance (GIDA), established in 2020, which is "a network of Indigenous researchers, data practitioners, and policy activists advocating for Indigenous Data Sovereignty within their nation-states and at an international level" (GIDA, n.d.-a). GIDA also works to advance the CARE Principles for Indigenous Data Governance (GIDA, n.d.-b);
- the First Nations Information Governance Centre (FNIGC), which was established in Canada in 2009 "with a special mandate from the Assembly of First Nations' Chiefs-in-Assembly (Resolution #48, December 2009)" (FNIGC, n.d.). FNIGC supports Indigenous data sovereignty and "the development of information governance and management at the community level through regional and national partnerships" (FNIGC, n.d.). The primary work of FNIGC involves data-gathering initiatives in First Nations communities. It has also developed the ownership, control, access, and possession (OCAP) standard for how to conduct research with First Nations; and
- the National Inuit Strategy on Research, developed by Inuit Tapiriit Kanatami in 2018, which advocates for advancing Inuit governance in research, including access, ownership, and control over Inuit data (ITK, n.d.).

Indigenous perspectives on AI can provide valuable insights that challenge culturally dominant assumptions and biases

In addition to working towards Indigenous data sovereignty, Indigenous researchers — such as those involved in the Indigenous Protocol and Artificial Intelligence Working Group — are studying the relationship between AI and Indigenous perspectives. In 2020, the Working Group published a position paper that examined a variety of Indigenous perspectives, knowledge systems, and technological practices related to AI (Lewis, 2020). Themes explored in the document include Indigenous control and sovereignty over hardware, software, and data; the role of Traditional Knowledge within AI systems and the need to protect such knowledge; designing AI systems in accordance with Indigenous ethical frameworks; the potential dangers that AI may pose to Indigenous communities "as an extension of colonial practices of exploitation, extraction and control;" the role of Indigenous languages in AI and the relationship between language and computational processes; and the nature of the relationship between human beings and AI (Lewis, 2020).

The Working Group focuses on how Indigenous epistemologies and ontologies can inform the development of AI. Such perspectives are intended to provide an alternative to the "Western rationalist epistemologies" that drive current AI development and the potential prejudices and biases such perspectives might contain (Lewis, 2020). Indigenous worldviews on AI can serve to "challenge the fundamental anthropocentrism of Western science and technology" by offering "epistemologies [that] refuse to center or elevate the human" in favour of technology development that is based on "principles and practices of social and environmental sustainability" (Lewis, 2020).

The CARE Principles for Indigenous Data Governance may help facilitate ethical use of Indigenous data

The CARE Principles for Indigenous Data Governance, which are an example of data stewardship and management principles that facilitate responsible and ethical data sharing and use (Section 3.1.4), were developed by the Indigenous Data Sovereignty Group within the RDA. These principles complement the existing FAIR (Findable, Accessible, Interoperable, Reusable) principles for scientific data management and stewardship and are meant to be implemented in tandem (Carroll *et al.*, 2020). The CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles are a direct response to the FAIR data governance principles — which may ignore power differentials and historical context — and are designed to acknowledge the role of data in advancing Indigenous innovation and self-determination (Carroll *et al.*, 2020; GIDA, n.d.-b). Importantly, the CARE principles do not only apply to Indigenous or Traditional Knowledge but also to scientific data (Carroll *et al.*, 2021).

5.4 EDI in AI Research and Access to AI Technology

Access to AI technology raises issues of EDI. Although limited data exist about the gender diversity of researchers in the field of AI, there is an overall lack of data regarding other demographic factors such as age, race/ethnicity, (dis)ability, sexual orientation, and socio-economic status (Young *et al.*, 2021). For a discussion on the intersection of EDI and AI in the research funding system, see Chapter 4.

There is an overall lack of gender and racial diversity in the field of AI research

Globally, in 2018, fewer than 22% of professionals in the fields of AI and data science were women; in Canada, the figure rises to approximately 24% (WEF, 2018).[13] Women in AI are more likely to hold a job with lower status and pay compared to men (e.g., "usually within analytics, data preparation and exploration, rather than the more prestigious jobs in engineering and machine learning"), are under-represented in industries with more technical skill demands, and have a higher turnover and attrition rate than men (Young *et al.*, 2021). Women authors account for only about 15% of AI papers

13 Note: this and several other studies referenced in this section do not distinguish between gender and sex and may use terms such as "woman" (gender) and "female" (sex) interchangeably. This report uses gender-based terminology in such cases.

published on arXiv (Element AI, 2020) and only about 18% of authors at leading AI conferences (Element AI, 2019). Similarly, only 16% of "tenure-track faculty whose primary research focus area is AI" are women (HAI, 2021),[14] and only 15% of AI research staff at Facebook and 10% at Google are women (Simonite, 2018). Women are also underrepresented in AI patenting: globally, there was one woman identified for every three men in AI patenting, while in Canada, the ratio decreased to one woman for every six men (ISED, 2019a). No reported data exist for transgender workers or other gender minorities (West *et al.*, 2019).

Although there are fewer data regarding race/ethnicity in AI, the available information indicates that the field is generally lacking in racial diversity. For example, racial and ethnic diversity among new AI PhDs residing in the United States skews approximately 46% White, 22% Asian, 3% Hispanic, 2% Black, and 2% multiracial (race and ethnicity information was unknown for 25%) (HAI, 2021). Although there are no data available about the racial/ethnic diversity of AI researchers across the private sector, only a small percentage of the total workforce at large tech companies is Black or Hispanic, including Google (2.5 and 3.6%), Facebook (4 and 5%), and Microsoft (4 and 6%) (West *et al.*, 2019). There are ongoing efforts to increase racial/ethnic diversity in the field of AI; for example, Black in AI is an organization that works to increase Black representation in the field. It has held multiple workshops at major AI conferences and has helped to significantly increase the number of Black participants at such conferences (HAI, 2021).

There are significant inequalities in access to AI and big data for research

There are currently high levels of inequality in the existing distribution of resources, infrastructure, and skills when it comes to the production, dissemination, and use of big data for scientific research (Leonelli, 2020). The digital divide traditionally refers to a gap in opportunity between "haves" and "have-nots" with respect to information and communication technology and the internet (Carter *et al.*, 2020). In the context of AI, this divide — which Yu (2020) terms the *algorithmic divide* — refers to differences among individuals, researchers, institutions, organizations, and geographical areas with respect to AI access, use, and outcomes. The algorithmic divide prevents a significant portion of the population from reaping the benefits of accessing AI technology, and it continues to widen (Leonelli, 2020; Yu, 2020). Furthermore, the divide affects the findings generated by data-driven research. For example, it limits the availability

14 A study by the World Economic Forum arrived at a somewhat different conclusion about gender diversity in AI in higher education, finding that it is one of the few industries globally in which women outnumber men in the AI talent pool (WEF, 2018). However, this study has been criticized by West *et al.* (2019) on methodological grounds related to its use of LinkedIn as the primary data source and its mechanism for attributing gender on the basis of first names.

of data about certain groups and geographical regions, and it affects which data are widely disseminated and which are kept confidential due to factors such as commercial value (Leonelli, 2020).

The divide is driven and shaped by social factors such as demographics, culture, and policy and regulation; technical factors such as infrastructure, algorithms, and training data; and socio-technical factors such as computer skills and literacy, as well as perceptions and beliefs about AI (Carter *et al.*, 2020). Economic factors may also exacerbate the algorithmic divide in AI for scientific research. For example, the high cost of computational resources and competition to attract AI research talent means that AI capability is often concentrated in private sector companies rather than universities or the public sector (OECD, 2018b). As stated by the OECD (2018a), "[t]his may lead to the concentration of scientific discovery and raises concerns about excessive monopoly of scientific knowledge." In Canada, public investments in the AI sector have been found to primarily benefit the private sector (Brandusescu *et al.*, 2021).

> "In Canada, public investments in the AI sector have been found to primarily benefit the private sector."

In Canada, the digital divide is most apparent in rural and northern areas of the country, which often have significant deficits in internet connectivity (CCA, 2021b). For example, fewer than half (48%) of people in Canada living outside large population centres (i.e., with a population greater than 10,000) have high-speed internet access (i.e., greater than 50 Mbps), compared with over three-quarters (76%) of those within such areas (StatCan, 2021b). Similarly, compared to the Canadian average (92%), lower rates of internet usage are found among Indigenous Peoples (88%), people with a disability (85%), unemployed people (85%), and people over 75 years of age (62%) (StatCan, 2021a). There is little to no publicly available information about the algorithmic divide in Canada, such as the proportion of public or private sector researchers with access to AI technology or their demographic profiles. Such information would be useful in developing policies to facilitate the use of AI for science and engineering in Canada.

The existing inequities in the AI research system may result in poorer research outcomes

As noted in Section 5.2, increasing diversity in the AI research system is important for identifying and mitigating potential discriminatory biases in AI. However, the value of diversity in AI extends beyond ethical and social concerns related to fairness; in addition, diversity in research utilizing AI may result in better epistemic outcomes for science and engineering. In a 2012 report, the

CCA's Expert Panel on Women in University Research found that "increasing the pool of Canada's researchers by opening opportunities to women and other underrepresented groups can generate stronger research outcomes" (CCA, 2012). Subsequent studies have found that researchers from groups that are under-represented in their discipline tend to produce higher rates of scientific novelty and new conceptual linkages (Hofstra *et al.*, 2020) and that gender diversity may lead to better science (Nielsen *et al.*, 2017).

Other studies have found that cognitive diversity — that is, diversity in perspectives or information-processing styles — is linked to better problem-solving in teams (Reynolds & Lewis, 2017). Cognitive diversity may also be epistemically beneficial for scientific research by spurring scientific creativity and leading researchers to seek out and pursue new approaches, research methods, types of evidence, hypotheses, and theories (Rolin, 2019). Moreover, many philosophers of science and social epistemologists have argued that, under the right conditions, increasing demographic diversity may concomitantly increase cognitive diversity and thus produce epistemic benefits for scientific research (Fehr, 2011; Rolin, 2019).

5.5 Labour Market Impacts

Although the labour market impacts of AI have been extensively studied (see OECD (2021b)), there is comparatively little information on how AI might affect employment in the fields of science and engineering. However, an analysis by Webb (2020) found that the occupations most exposed to displacement by AI include metallurgical and materials engineers, chemical engineers, physicists and astronomers, atmospheric and space scientists, engineering and science technicians, and other science and engineering professionals. AI is likely to affect the labour market differently in different fields of science and engineering. For example, fields of research that involve more manual activities, such as animal science and archaeology, may be less exposed to AI (Brynjolfsson *et al.*, 2018), as are fields of research that involve reasoning about novel situations (Webb, 2020).

The degree to which AI would impact the science and engineering labour markets in Canada is unclear. In 2016, the size of the labour force in the National Occupational Classification category "natural and applied sciences and related occupations" numbered over 1.27 million people, or nearly 7% of the total labour force. Approximately 725,000 of those individuals were employed in "professional occupations in natural and applied sciences," whereas 549,000 were employed in "technical occupations related to natural and applied sciences" (StatCan, 2016). AI might also affect jobs downstream of science and engineering research, such as those involved in commercialization of new discoveries. However, there is very little evidence about how AI will impact these occupations. The use of AI in science and engineering may indirectly create new jobs in these areas via its potential to produce innovations and scientific breakthroughs (Lane & Saint-Martin, 2021).

AI may displace some science and engineering occupations, but it is likely to transform many more

Although AI will likely disrupt the labour market in science and engineering, it may be more likely that AI will transform many occupations rather than displace them (Lane & Saint-Martin, 2021). Brynjolfsson *et al.* (2018) found the vast majority of occupations have tasks that are both highly suited and poorly suited to replacement by AI and argue that debates about the impact of AI on labour markets should focus on the redesign, rather than replacement, of jobs due to AI. In the context of science and engineering, the practical impacts of AI on how research is carried out may require professionals in these fields to adopt new mindsets that reconceptualize the value that they bring. As one chief technology officer of a chemical manufacturer using AI for industrial biotechnology research put it, "[i]f you see yourself as just somebody setting up a machine, you may well end up missing the big picture" (Mullin, 2021).

"In the context of science and engineering, the practical impacts of AI on how research is carried out may require professionals in these fields to adopt new mindsets that reconceptualize the value that they bring."

Furthermore, despite the potential for AI in generating hypotheses and guiding research priorities (Section 4.1), some argue that scientists and other kinds of innovators will still be required to identify and prioritize new problems or opportunities (Lane & Saint-Martin, 2021).

To facilitate building human capacity for the AI labour market, policies will be needed in areas of education and training, attracting and retaining AI talent, and empowering people to effectively use and interact with AI systems (OECD, 2021b). Countries are launching a range of policy initiatives, such as formal education programs, vocational training and lifelong learning programs, financial and non-financial support to attract and retain AI talent, and the development of academic partnerships between public and private AI research institutions (OECD, 2021b). In addition, several countries, including France, Germany, Czechia, and Poland, all have national institutions that closely monitor the impact of AI on the labour market. Between 2015 and 2019, Canada was one of the top destinations for the migration of AI talent, behind Luxembourg, the United Arab Emirates, and Ireland (OECD, 2021b).

Adapting the workforce to the impacts of AI will also require reforms to educational curricula and professional standards and codes (Villeneuve *et al.*, 2019), the development of new labour standards and labour agreements, and addressing AI-related EDI issues (OECD, 2021b). EDI issues will require particular attention given the current general lack of diversity in AI research and the existing inequalities in access to AI technology (Section 5.4).

5.6 Environmental Impacts of AI Systems

AI systems can have a significant environmental impact due to the energy required to power computational infrastructure. For example, an analysis by Strubell *et al.* (2019) found that the development and training of certain common AI models for natural language processing can produce up to 284 tonnes of carbon emissions, equivalent to the average lifetime emissions of five cars. However, Patterson *et al.* (2021) dispute this estimate, finding that the emissions of such models are only about 5.3% of what Strubell *et al.* (2019) claim.[15]

Lacoste *et al.* (2019) have developed a carbon footprint calculator for machine learning that allows researchers to calculate the environmental impact of their work (Schmidt *et al.*, n.d.). However, post-hoc estimates of carbon emissions are often less accurate than measuring the actual energy usage of AI systems (Patterson *et al.*, 2021). Tools such as CodeCarbon and CarbonTracker offer code that can be integrated into AI systems to provide estimates of the amount of CO_2 emissions they generate (Anthony *et al.*, 2020; CodeCarbon, 2021).

Factors that influence the carbon emissions of AI include the geographical location and type of the computational infrastructure on which the AI model is trained, the type of energy grid being used, and the length of the training process (Lacoste *et al.*, 2019). Some of the energy used to power AI may be derived from renewable sources or offset through the use of carbon credits (Strubell *et al.*, 2019). However, it is unclear whether buying carbon credits is effective in reducing overall energy usage. Furthermore, renewable energy is not always available; in Canada, about 67% of electricity was generated from renewable sources (primarily hydroelectric) in 2018, although this ratio varies significantly among provinces and territories (NRCan, 2020). Cloud-based data centres that are optimized to take advantage of renewable energy are useful for reducing the emissions associated with AI models (Patterson *et al.*, 2021). Patterson *et al.* (2021) provide several examples of techniques that can improve the energy efficiency of machine learning models without a loss of accuracy.

15 For clarity, Patterson *et al.* (2021) claim that the estimate by Strubell *et al.* (2019) is 18.7 times too high, which is equivalent to the revised estimate of approximately 5.3%.

Several researchers (e.g., Strubell *et al.*, 2019; Henderson *et al.*, 2020; Schwartz *et al.*, 2020; Patterson *et al.*, 2021) have argued that energy efficiency should be an evaluation criterion for publishing AI papers, alongside other criteria such as accuracy. Furthermore, Gupta *et al.* (2020) argue for the need for certifications for the environmental and social impacts of AI systems so that consumers and investors can make informed decisions. Taddeo *et al.* (2021) have developed 14 policy recommendations for environmentally sustainable AI; although many of the recommendations are specific to the EU policy context, "they are not meant to be EU centric but to contribute to the international debate from a European perspective."

The environmental impacts of AI are not limited to the GHG emissions associated with model development and training

The GHG emissions associated with the training, testing, and use of AI models represent only a small part of the total environmental impacts of AI. These also include the embodied carbon within AI systems, as well as other environmental impacts caused by the mining and mineral extraction and manufacturing processes associated with the production of AI components such as semiconductors (Mulligan & Elaluf-Calderwood, 2021). Furthermore, these impacts increase when considering the full life cycle of AI systems across the entire supply chain, which includes raw material extraction and manufacturing of components; the transportation of materials and components; the construction and installation of AI infrastructure; maintenance, repair, refurbishment, and upgrades to the system; and the end-of-life stage, including transportation, waste processing, and disposal (Mulligan & Elaluf-Calderwood, 2021).

Despite its environmental impacts, research that uses AI can also play a significant role in providing solutions to climate change, such as generating designs for smart grids and low-emissions infrastructure, and modelling climate change predictions (Dhar, 2020). Indeed, one recent analysis of research funding programs in the EU found 122 projects that use AI to address various aspects of climate change (Cowls *et al.*, 2021).

5.7 AI Security

AI-powered research is likely to have a wide range of implications for society, some of which are described above. However, the risk of social harms arising from such research could be exacerbated by security vulnerabilities in AI systems. Like any computer system, AI is vulnerable to cyberattacks. Such attacks may be difficult to detect due to the black box nature of certain types of AI models (Xue *et al.*, 2020) (Section 3.1.2). Attackers may target the training data, the AI model itself, or the underlying hardware or software infrastructure (AI HLEG, 2019). Attacks may have several different goals, including skewing or falsifying training data to reduce the accuracy of the model, causing the system to malfunction or make a mistake, or stealing AI models or sensitive training data (Xue *et al.*, 2020).

There is little research that specifically examines the role of cybersecurity in the context of AI for science and engineering. Nevertheless, issues that arise in this context include IP protection, privacy protection, and scientific accuracy. Attacks that cause an AI model to generate inaccurate results could have serious negative social impacts if such results are used in a way that exacerbates societal bias or discrimination, or even affects public safety, in the case of engineering designs that are incorrectly assessed as safe. Articles 5.3 and 5.4 of TCPS 2 require both researchers and institutions or organizations to establish security safeguards that cover the full life cycle of data gathered and used in research involving human subjects (CIHR *et al.*, 2018b).

> "Attacks that cause an AI model to generate inaccurate results could have serious negative social impacts if such results are used in a way that exacerbates societal bias or discrimination, or even affects public safety, in the case of engineering designs that are incorrectly assessed as safe."

To proactively defend against attacks, it is necessary to undertake security evaluations of AI systems at the design stage, an approach that Xue *et al.* (2020) call *design-for-security*. This involves subjecting the system to strong mock attacks, such as those in which it is assumed that the attacker has full knowledge about the model, data, and defence techniques, as well as the ability to manipulate the model (Xue *et al.*, 2020). In order to help AI developers ensure the security of their systems, several researchers have developed an open-source library of adversarial example attacks that can be used to test the robustness of AI models (CleverHans, 2021).

Both the European Commission (2019) and the OECD (2021a) identify security and resilience to attack as a requirement for trustworthy AI. However, compared to other countries, Canada's AI policies and strategies have not devoted much attention to this issue (Cussins Newman, 2019). Several countries, including Canada, have emphasized the need for common standards to help address security issues related to AI (OECD, 2021a). Such standards are being developed by organizations such as the Institute of Electrical and Electronics Engineers and the British Standards Institute (Muller, 2020). In addition, the Danish government, in collaboration with industry stakeholders, has developed a certification scheme known as the Joint Cybersecurity and Data Ethics Seal that is granted to companies that meet ethical and cybersecurity requirements for AI-related data (OECD, 2021b).

Security risks to AI systems are prevalent and multifaceted, with governments and industry stakeholders seeking to reduce the probability and impacts of IP theft

The development of AI is increasingly viewed as a global race, with many nations hoping to harness these technologies and realize significant economic returns. AI systems and tools are already being applied for military purposes, lending an additional geopolitical dimension to AI R&D (Boulanin *et al.*, 2020). The heightening global competition and high stakes involved raise security concerns.

Although several of the above security risks lie outside of the scope of the Panel's charge, some have legal implications in science and engineering for governments and developers to consider. Chiefly, the valuable IP produced through AI R&D is susceptible to theft or espionage (Friesen, 2021). The intangible nature of algorithms and data, combined with the fact that AI systems can be repurposed from scientific applications to other applications (Somers, 2021), has prompted several measures on the part of governments in Canada and abroad to address this vulnerability. In 2021, the Canadian federal government launched initiatives to mitigate some of these risks by defining new guidelines for research partnerships based on national security (ISED, 2021b) and national security risk assessments for university researchers seeking federal funds (Fife & Chase, 2021). AI has also been identified as a "sensitive technology," wherein greater scrutiny will be applied to commercial activities involving foreign investors in the recently updated Guidelines on the National Security Review of Investments (ISED, 2021c). These efforts accompany other legal developments, such as those in the United States and in provisions surrounding IP in major international trade treaties, whereby penalties for the theft of trade secrets, in particular, are being made more stringent (Ciuriak & Ptashkina, 2021). Recent reforms to the criminal code to implement requirements agreed to in the *Canada-United States-Mexico Agreement* (CUSMA) have elevated trade secret theft to a criminal act, reflecting government concerns (Parliament of Canada, 2020). Existing cybersecurity legislation can apply to such situations and treat them as hacking, but these laws are seen by some to be outdated (Brodt *et al.*, 2021). Other legal, regulatory, and policy issues related to AI are explored further in Chapter 6.

6

Implications of AI for Canadian Law, Regulation, and Policy

Chapter Findings

- Access to high-quality data is crucial for the development of AI for science and engineering. Governments have established open data policies, but some data requires additional governance to protect privacy, ensure ethical reuse, or otherwise protect the public interest.
- Data can be valuable assets, and tensions exist between enabling access and maintaining commercial advantages. This leads to concerns that high-value data will be kept confidential.
- Obtaining IP protection for several constituent elements of AI systems is a challenge, and discoveries or designs produced without humans might be ineligible for protection under current laws.
- AI systems challenge conventional legal liability frameworks, given their potential lack of transparency, uncertain security, and challenges in establishing causation or allocating responsibility. The resolution of these issues may vary by jurisdiction.
- Despite initiatives to govern AI in the public sector, Canada's regulatory approach to AI governance will require harmonized federal and provincial/territorial action. Regulatory developments in partner countries may influence the course of reforms in Canada.

There is a global race to develop AI techniques and technologies for both economic and societal benefit. The emerging legal and policy environment is uncertain and lacks harmonization, presenting challenges to resolve at all stages of the AI development life cycle. In the early stages, the development of AI hinges on data — their creation, storage, acquisition, use, and protection. At the commercialization stage, success requires the effective management of IP which is not necessarily straightforward with AI. This challenge affects both AI systems and their outputs, with additional complications foreseen as AI systems take on more active roles in creating IP.

6.1 Access to Data

A common feature of all AI systems is the need for data. AI must generally be trained on a large volume of high-quality data to achieve accuracy and mitigate possible biases. From a practical standpoint, AI developers and users may not possess all the data they require in-house and may seek to access additional external sources. The data used for the purposes of developing AI systems will vary widely in type, in form, and in origin. Accompanying this variability of data sources are distinct governance mechanisms. Personal data are subject to a number of different provincial/territorial and federal laws (OPC, 2021a), while research data are governed according to formal and informal policies dictated by numerous stakeholders, such as funding agencies and organizations where R&D takes place.

> "The importance and value of data for AI R&D, together with the varying priorities among stakeholders, heighten the importance of controlling data use and data flows using legal means, policies, and data infrastructure."

For research purposes, AI systems will invariably draw on the vast amounts of data made available openly or through licences from repositories. However, some stakeholders may wish to limit or restrict access to certain data or protect them through legal means, such as copyright or trade secret law. The importance and value of data for AI R&D, together with the varying priorities among stakeholders, heighten the importance of controlling data use and data flows using legal means, policies, and data infrastructure. Contracts offer private actors a way to set terms and conditions for data access and use. For example, Facebook was able to include its users in an experiment without explicit consent via its user agreement (Box 3.1). Other organizations may impose data-sharing obligations via contractual conditions, allowing them to effectively mandate open data.

Governments and other stakeholders seek to promote AI by providing open access to data; however, challenges remain regarding interoperability and privacy

Governments are increasingly aware of the value of the data that they hold for AI and innovation (see The Royal Society (2017)). The EU has identified public sector data as a pillar of its digital economy and in 2019 created a legal framework for high-value datasets, which can be accessed openly and reused with AI development in mind (European Parliament, 2019b). In recent years, Canada has taken steps to be a leader in open government data (Web Foundation, 2018), with initiatives such as its Open Data Portal and an open-by-default policy (GC, 2021b, 2021g). The Canadian Digital Charter also emphasizes interoperable data frameworks and "Open and Modern Digital Government" among its core principles (ISED, 2019b). The government aims to provide open access to all federally funded science data in a FAIR-compliant format by 2025 (GC, 2020c).[16]

16 The federal government has expressed its intention to extend open science policies beyond research carried out by government agencies to research funded by the federal government; however, details and timelines for implementation were not available at the time of writing (GC, 2020c).

Open data initiatives also exist at other levels of government. For example, the Quebec government partners with municipal government agencies and certain non-governmental organizations for its open data portal (Données Québec, 2021). Governments are also creating new structures and roles for overseeing the sharing and stewardship of data held by government agencies and departments. In Canada, the 2021 federal budget included funding for a new Data Commissioner, who might play a role in overseeing the availability of and access to high-quality public data, similar to New Zealand's Chief Data Steward or Australia's Data Commissioner (New Zealand Government, 2020; Australian Government, 2021). At the provincial/territorial level, Ontario has had a Chief Digital and Data Officer since 2017 (Gov. of ON, 2017) and is currently establishing the Ontario Data Authority to promote access to data (Gov. of ON, 2021b).

> "Government efforts to improve access to data for the potential benefit of society are not limited to dissemination, but also involve the collection of data and the establishment of data infrastructure."

Government efforts to improve access to data for the potential benefit of society are not limited to dissemination, but also involve the collection of data and the establishment of data infrastructure. Data infrastructure is not only physical; it also comprises datasets and identifiers, standards, policies, organizations governing infrastructure, and user communities (Dodds & Wells, 2019). Canada's Geospatial Data Infrastructure (CGDI) represents an example of an end-to-end platform for collecting and sharing spatial data (GC, 2020b). By monitoring geographically relevant features and boundaries in Canada, CGDI can inform a wide range of activities — from natural resource management to Indigenous land-claim settlements — and represents a versatile digital public asset relevant for many stakeholders and users (NRCan, 2019). MacGregor (2018) suggests that the expansion of data-collection activities of this type — for example, directing public investments to update existing sensor infrastructure for modern digital communications capabilities — offers the potential to transform Canada's primary industries.

Despite the value of open government data, risks and limitations exist. As the quantity of data grows, so do the requirements for expertise and infrastructure to make use of it, a trend that favours large, established institutional actors (Davies *et al.*, 2019). In the case of spatial data, as in the CGDI example, the provision and use of data can be geographically heterogeneous (Johnson *et al.*, 2017). A recent report on CGDI user needs echoes these challenges and identifies the need for additional efforts in order for Indigenous communities to extract greater value from the use of CGDI, such as providing resources in Indigenous languages (NRCan, 2019).

Other tensions may arise due to the sensitive nature of some data. The Ontario Health Data Platform was launched in the context of the provincial response to the COVID-19 pandemic and aims to provide greater access to large collections of integrated health data (OHDP, 2020). The use of AI in healthcare is beyond this Panel's mandate; however, obtaining representative health datasets is important for improving the accuracy of AI models, such as those used in medical devices (CIFAR, 2021).[17] Providing that access carries the risk of personal health data being used without consent, or infringement of privacy rights (Reznick *et al.*, 2020). Such access must be provided in a way that is consistent with laws protecting personal health information. The Ontario Health Data Platform requires that researchers register and agree to terms of use in an attempt to manage the privacy risks that exist when repositories are used to share sensitive and/or personal data (Wylie, 2018; Scassa, 2019b). Examples such as the Ontario Health Data Platform, as well as provisions to facilitate access to personal data for research purposes in proposed reforms to privacy legislation (Section 6.4), highlight the willingness of governments to provide greater access to personal data in support of the AI sector. This trend raises the importance of developing technological and legal data protection mechanisms to accompany open data initiatives.

Text and data mining and data scraping can be essential tools for developing AI but can also potentially be unlawful; several jurisdictions are relaxing copyright protections to eliminate this barrier

Data extracted through text and data mining (TDM) techniques, such as text, images, and audio recordings, can be rich sources of data for AI. However, even where researchers have permission to access copyright-protected works (e.g., licenced access to a database of materials), they may not have a right to reproduce, modify, or use these works in TDM because of copyright protection (Flynn *et al.*, 2020). Some academic publishers use licensing agreements to control access to their works for TDM applications (Caroll, 2019) or to the data accompanying articles published in their journals (Baker *et al.*, 2019). Since TDM requires the creation of local copies of data or the conversion of material into machine-readable formats (Flynn *et al.*, 2020), it could therefore infringe on the economic rights of copyright holders (Craig, 2021).[18] Data scraping, where software tools are used to extract data from web pages or online databases,

17 Additional discussions of the opportunities and risks of integrating AI into Canadian healthcare can be found in Reznick *et al.* (2020).

18 "Fair dealing" exceptions in copyright law might allow for text and data mining where the research activities do not damage or limit the "market" for a copyright holder (Craig, 2021), but entails a case-by-case analysis and leaves considerable uncertainty as to the legitimacy of some uses.

raises similar concerns. Not only does scraping risk infringing copyright through the creation of potentially unauthorized copies, but it may also violate the terms and conditions of websites (Scassa, 2021a).

In some jurisdictions, including the United Kingdom and the EU, explicit copyright exceptions have been introduced for certain TDM applications (Kelly, 2016; European Parliament, 2019a). These exceptions provide indirect support for nascent AI industries by eliminating risks of infringement. Proponents of TDM exceptions argue that they support scientific research and that they can level the playing field for access to data while remaining consistent with the objectives of copyright law (Craig, 2021). Detractors, however, argue that the authors of the mined works will not be fairly rewarded through these exemptions (INDU, 2019) because, in some cases, the data contained in the works might require a significant investment to create (Mercurio & Yu, 2021).

Not all TDM exceptions have the same scope, creating the potential for conflict across jurisdictions. In the United Kingdom, TDM exceptions are restricted to non-commercial activities (Kelly, 2016). The EU directive, meanwhile, restricts exceptions to specific types of users, such as research organizations (European Parliament, 2019a). Private-public partnerships (PPPs) that exist in the AI domain may be excluded from TDM exceptions due to the restrictions placed on commercial activities (Flynn *et al.*, 2020). Though the EU specifically considers PPPs within the scope of its exception framework (European Parliament, 2019a), the restriction on TDM for commercial activities is seen as a "murky area" in the United Kingdom (The Royal Society, 2017). A similar lack of harmonization exists for data scraping, with diverging approaches in the EU and the United States (Scassa, 2021a). Canadian entities in international collaborations with academic, public, or private partners will need to be aware of the legality of any TDM and scraping applications they pursue, both in Canada and in the jurisdictions where their partners are based. In Canada, copyright exceptions for TDM have not yet been implemented. Following stakeholder consultations, however, the House of Commons Standing Committee on Industry and Technology recommended amendments to the *Copyright Act* to provide explicit exemptions for "informational analysis" to permit TDM activities without the need for a licence (INDU, 2019). In 2021, to build on the evidence gleaned from the parliamentary review, Innovation, Science and Economic Development Canada launched a consultation on a Modern Copyright Framework for AI and the Internet of Things (ISED, 2021a).

Obtaining IP protection for data can be difficult unless it is kept confidential as a trade secret

The high value of data for R&D in AI leads to several tensions. On the one hand, transparency and openness are increasingly being encouraged to promote trust and facilitate progress, particularly in academia and in government. On the other hand, data are valuable, and protection may be desired to maintain commercial advantage, although data are not necessarily easily protected as IP. Works that can be mined for data may be protected by copyright, but data themselves are generally only protected by copyright law to the extent that they form part of an original compilation of data (Scassa, 2019a). Even then, only the original selection or arrangement of the data — and not the underlying data — is protected (Craig, 2021).[19] As such, one of the more valuable constituent elements of an AI system — the data used to develop it — may either be ineligible for copyright protection or else only weakly protected (Medeiros *et al.*, 2021). Although the EU offers *sui generis* protection for databases, such protection has important limitations in the contemporary data context because exceptions exist for the extraction of data from protected databases for scientific research (European Commission, 2018).

"One of the more valuable constituent elements of an AI system — the data used to develop it — may either be ineligible for copyright protection or else only weakly protected."

This environment contributes to the use of trade secret law to protect data. Trade secret protection requires that data or information remain secret; that the data have commercial value because they are secret; and that these data have been subject to reasonable measures, under the circumstances, to protect their secrecy (WTO, 2017). Maintaining the confidentiality of trade secrets is essential to their continued existence (Malone, 2020). Unlike copyright or patent protection, trade secrets may, in theory, be protected in perpetuity. Protection is lost when confidential data fall into the public domain (Scassa, 2021b). Medeiros *et al.* (2021) suggest that trade secrets represent an important layer of IP strategy for nascent AI technology companies, particularly during the development phase. In Canada, however, trade secrets are generally protected by provincial/territorial law rather than federal law (Scassa, 2021b). Courts in different provinces or territories may pay greater attention to the data, or to how the data were collected or produced, to determine their eligibility for protection (Malone, 2020).

19 The issue of whether data may be "original" remains unclear and could cause copyright law to evolve in this area.

Although trade secrets may be appealing to some innovators for protecting data, their effectiveness remains to be tested in the context of AI in Canada (Malone, 2020). Scassa (2021c) and Malone (2020) both argue that it is possible for public interest to override the confidentiality of data in some circumstances. Mechanisms for doing so exist at both federal and provincial/territorial levels — such as access to information laws (GC, 1985b; MGCS, 2021) — and have been used to obtain disclosure of clinical trial data held by Health Canada. This gives rise to tensions with respect to commitments made in international trade agreements (Malone, 2020). Trade secrecy also increases the lack of transparency characteristic of black box AI systems (Section 3.1.1). As Citron and Pasquale (2014) suggest, by preventing others from understanding how and why automated systems make decisions — for example, in the context of credit scoring —trade secret protection of data supports a "black box society" in which the allocation of benefits or opportunities is opaque to those affected.

International data flows are a growing consideration in the globally interconnected digital economy, and AI practitioners and collaborators will face jurisdictional variability in the local governance and protection of data

Data flows across borders are a growing issue because of the integration of digital data into value chains of commercial activities at international scales (Lopez-Gonzalez *et al.*, 2021). The borderless character of digital data exacerbates challenges in data governance because regulations, data protection, and IP laws are territorial and therefore will be different from one jurisdiction to another (Lopez-Gonzalez *et al.*, 2021).[20] By stipulating which laws govern an agreement, governing law clauses may be a mechanism of certainty in this regard. Nevertheless, jurisdictional issues complicate the development of AI systems for scientists and engineers, who may require access to data that are unavailable to them within their jurisdiction. For example, cross-border data flow issues may be unavoidable for research — especially research collaborations— because of different approaches to TDM in domestic copyright laws or incompatibilities in other areas of data governance (Flynn *et al.*, 2020).

Modern trade agreements commonly include provisions that aim to facilitate data flows, with some more recent examples featuring binding provisions and enforcement mechanisms (Lopez-Gonzalez *et al.*, 2021), such as the *Comprehensive and Progressive Agreement for Trans-Pacific Partnership* (CPTPP) and CUSMA (Leblond, 2019). "Mega-regional" trade deals of this type have come to shape the landscape for data flows across borders instead of agreements made by

20 Consider, for example, the difficulty of enforcing Canadian privacy laws against Clearview AI for its image scraping activity (see Box 3.2).

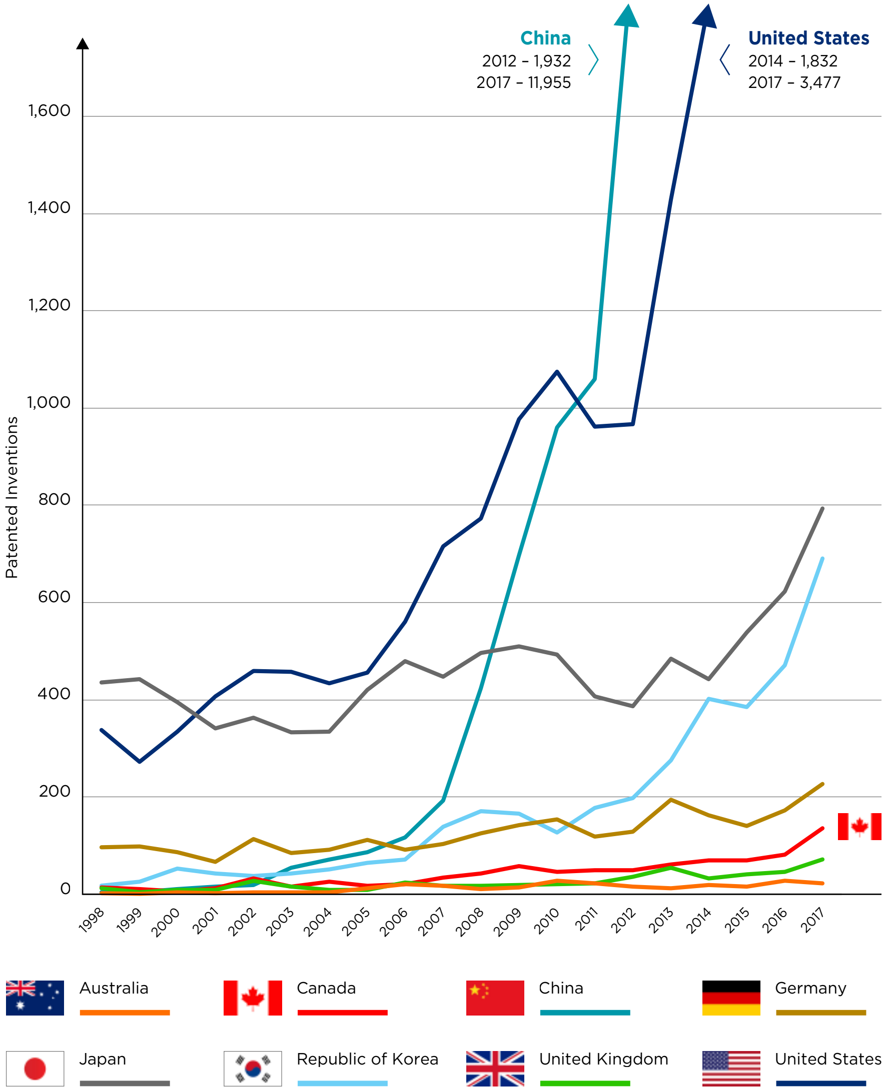
multilateral institutions such as the World Intellectual Property Organization (WIPO), which governs other forms of IP (de Beer, 2020). Flynn *et al.* (2020) argue that WIPO should nevertheless define norms with respect to cross-border issues, given its role in defining international IP policy. Conflicts may exist between internal regulatory or policy positions and the requirements agreed upon by Canadian trade officials and their international partners. For example, the CPTPP contains provisions misaligned with requirements in Canadian regulations governing data localization (Leblond, 2019), and the operationalization of the Digital Charter could be constrained by provisions in CUSMA (de Beer, 2020). SMEs in Canada will need to navigate these cross-border IP issues relating to data to access larger markets.

6.2 Commercialization of Machine-Generated Outputs

AI R&D for science and engineering may result in IP-protectable outputs that offer the potential for economic and other societal benefits. The outputs of such AI systems, as well as their methods, algorithms, data, and intermediate discoveries, can each represent potentially valuable forms of IP. The effective management of this IP is a key to Canada's performance in the intangible innovation landscape (Expert Panel on Intellectual Property, 2020; Lamb & Munro, 2020). For policy-makers at provincial/territorial and federal levels, deriving societal benefit through IP commercialization is a priority due to the large public investments in AI. This pressure is conceivably heightened by persistent observations — not limited to AI — that IP produced in Canada is poorly exploited (CCA, 2013, 2018; Hinton, 2020). In this highly competitive field, innovators in Canada will need to overcome existing barriers and determine approaches to resolve emerging questions surrounding IP generated by AI.

Patenting activity in AI has accelerated dramatically in recent years, complicating the landscape for innovators and renewing calls for capacity building in IP education and management

Patents temporarily provide innovators with protected market access in exchange for disclosing an invention. They are also an important vehicle for protecting and commercializing IP. A patent application must provide sufficient disclosure of the invention (GC, 1985a), such that a person skilled in that field could reproduce it (CIPO, 2021). There is a global race to patent IP in the AI domain, with implications for researchers and entrepreneurs in Canada (Expert Panel on Intellectual Property, 2020). Figure 6.1 reveals an increasing rate in patenting since the late 2000s, but with differences from country to country. These data — based on the location of the patent's assignee(s) and not necessarily of its inventor(s) — show that the greatest increases in patenting have occurred in the United States and China, and more recently in Japan and Korea (ISED, 2019a).



Reproduced with the permission of the Minister of Industry, 2022 (ISED, 2019a)

Figure 6.1 AI Patenting Activity in Select Countries, 1998–2017

Trends in the number of patented AI inventions on a worldwide scale over the previous two decades.

The main subject areas within AI where patenting activity has been the strongest have also varied and diversified over the same period, reflecting rapid advancements in the field. The top keywords and International Patent Classification categories have changed over the past decade, with machine learning becoming dominant (ISED, 2019a; Habibollahi Najaf Abadi & Pecht, 2020). However, as many as 10% of AI patents are not classifiable under currently used sub-categories (ISED, 2019a). This demonstrates how the rapid pace of technological evolution and the general-purpose nature of AI may complicate efforts in monitoring trends within AI despite the development and refinement of categorization frameworks by the WIPO, the OECD, and various other national organizations focusing on IP. Both the global nature of AI development and the fact that the novelty of inventions can defy categorization contribute to a complex IP landscape that entrepreneurs in this field must navigate to operate both domestically and in foreign markets.

"Both the global nature of AI development and the fact that the novelty of inventions can defy categorization contribute to a complex IP landscape that entrepreneurs in this field must navigate to operate both domestically and in foreign markets."

Non-trivial jurisdictional issues add to this complexity because the requirements for patentability may vary from one country to another (ISED, 2019a). For example, software in the form of computer code can be patented in the United States; whereas in the EU and Canada, a computer program must be considered a computer-implemented invention to be patentable (WIPO, 2020; CIPO, 2021). Jurisdictional issues are exacerbated in certain technological areas with particularly complex networks of IP rights-holders, such as biotechnology or telecommunications. This may lead to "patent thickets," which are large clusters of patents over key technologies in the field. Those who seek to innovate in this area must obtain multiple licences, which in turn creates a considerable barrier to access and innovation. Patent thickets also incentivize defensive patenting, whereby patents are filed and collected to establish a portfolio that can subsequently be used to impede competing firms or as a bargaining tool (Gallini & Hollis, 2019).

Avoiding litigation and identifying patenting opportunities or strategies requires an understanding of issues relating to multiple jurisdictions and established IP landscapes. Innovators in Canada may lack the IP capacity or the access to IP expertise needed to address these issues (Gallini & Hollis, 2019). Patent drafting in this area requires costly expertise (Expert Panel on Intellectual Property, 2020). Only 2% of SMEs in Canada held a patent in 2019 (CIPO, 2019). Moreover, there is heterogeneity in IP governance policies at Canadian research institutions that may have divergent policies around the ownership of inventions. There is a debate among stakeholders as to whether different institutional approaches offer flexibility to researchers or confusing barriers for industry partners (Expert Panel on Intellectual Property, 2020). Institutions and individual researchers may value IP differently and may wish to pursue distinct strategies. Recent reports point to opportunities to build capacity in Canada by bolstering IP education (Gallini & Hollis, 2019) and leveraging IP expertise that may be concentrated in specific geographic areas (Expert Panel on Intellectual Property, 2020). The Intellectual Property Strategy launched by the federal government in 2019 aims to provide tools to help innovators in Canada manage these assets through IP legal clinics and patent pools, among other initiatives (GC, 2019), but the effectiveness of these policies is yet to be determined.

Patent protection is designed to incentivize the disclosure of inventions, but the applicability of this instrument to inventions made by AI systems is under debate

AI systems are already capable of wholly or partially producing what would conventionally be viewed as IP-protectable outputs. However, issues surrounding both patentability and copyrightability have been identified by WIPO in a discussion brief on IP and AI, reflecting the lack of international consensus (WIPO, 2020). Fundamentally, the spirit of patent protection is to define an incentive structure for innovation and to facilitate the disclosure and subsequent diffusion of inventions resulting from human ingenuity. A machine requires no incentive to innovate; however, the researchers who developed the machine may respond to incentives to stimulate ongoing innovation. As AI systems become more complex, particularly with machine learning, questions will arise regarding at what point something produced by that machine can no longer be attributed to a human inventor. In such cases, should that invention no longer be eligible for patent protection? To address this problem, some approaches propose either generating new legal frameworks for IP rights in AI or handling machine-generated IP using modified versions of existing frameworks, such as those pertaining to computer-implemented inventions (WIPO, 2020).

"As AI systems become more complex, particularly with machine learning, questions will arise regarding at what point something produced by that machine can no longer be attributed to a human inventor."

In patent law, it is the inventor who applies for a patent, raising the question of whether a machine can be an "inventor." According to Hagen (2021), the answer depends on a machine's capacity to intentionally conceive an invention by exerting its "mental processes." Some jurisdictions have been proactive in denying inventorship status to machines, with the European Patent Office (EPO) resolving in early 2020 to not grant patents for applications where an AI system is listed as the main inventor (Hervey, 2020). Other jurisdictions have recently opted to grant patents to machine inventors (Box 6.1) — first in South Africa and then in Australia, with the latter decision being justified as "consistent in promoting innovation" (IPWatchdog, 2021; Taylor, 2021). Inventions produced jointly by humans and AI systems remain patentable; however, the necessary contribution that a human should provide to qualify as an inventor is undefined (Hervey, 2020). There is a subsequent risk that inventorship or ownership status might be claimed in patent applications by humans who did not contribute to IP creation in order to circumvent contribution requirements and assert IP protection (WIPO, 2020).

Box 6.1 "DABUS," or the Artificial Inventor

An AI system dubbed "DABUS" (Device for Autonomous Bootstrapping of Unified Sentience) offers an example of patenting issues with respect to machine-generated IP. DABUS uses reinforcement learning to produce ideas that might qualify as inventions. The inventions the system has produced thus far have resulted in patent applications for a beverage container concept and for a method to modulate light sources in an attention-grabbing manner (The Artificial Inventor Project, 2021). Patent applications naming the machine as an inventor have been rejected in numerous jurisdictions, including the United Kingdom, the United States, the EU, and Taiwan, but were granted in South Africa and Australia (following appeal) (Chang, 2021; Egbuonu, 2021).[21] Although DABUS is not trained to solve any specific problem in a given topic area (Morrison, 2019), detractors of the decision to grant a patent and inventorship to DABUS cite the risks of AI systems being used as "patent thicket generators" to rapidly produce a large volume of patents in targeted areas to ward off competitors (Taylor, 2021).

21 The applicants have been appealing negative decisions. However, courts in the United Kingdom have rejected an appeal of its earlier decision not to grant a patent for this application (Bond, 2021).

The eligibility and appropriateness of copyright protection for machine-generated creations are uncertain due to incompatibilities with current legal frameworks

An important debate related to copyright law is whether machine-generated works should be protected, and if so, how authorship should be determined. In Canada, works produced by "authors" are protected by copyright. Craig (2021) suggests that, although the *Copyright Act* does not explicitly define authors as humans, it effectively implies as much through its requirements. For example, eligible authors must possess residency status in a given geographical location, and the copyright duration is linked to the lifespan of the author. The anthropocentric nature of copyright legislation poses additional problems in ascertaining the eligibility of content for protection. The subjective threshold regarding the "originality" required for a creation to qualify for copyright protection is not easily applied to machine creations (Craig, 2021).

Debates surrounding the appropriateness of copyright for machine-generated artistic works are ongoing (Deltorn, 2017; Craig & Kerr, 2021), and parallels can be drawn to creations in the context of science and engineering. Whether there will be reforms to copyright protection in Canada remains uncertain, but the federal government has begun a public consultation process to modernize its copyright framework with AI explicitly in mind (GC, 2021a). Craig (2021) argues that, on the basis of the *Copyright Act* in Canada and the history of copyright, the absence of copyright protection for machine-generated creations would not result in "underproduction" of original works because such protection does not incentivize machines. In the Canadian context, a standalone process that encourages the production of machine-generated artifacts could conceivably be created to offer limited protections as an analogue to copyright (rights, terms, limited monopoly), but would need to be technologically neutral to avoid quickly becoming obsolete (Craig, 2021). Alternatively, AI outputs that reach the "originality" threshold to qualify for copyright could simply exist in the public domain by default (Craig, 2021). The rate at which machine-generated content can potentially be created is so rapid that, if this content were protected, it would be challenging for human creators to compete and avoid infringement (Asay, 2020). Keeping machine-generated outputs in the public domain avoids this challenge and instead allows that content to inspire human creators (Craig, 2021).

Uncertainty regarding the patentability of machine-generated inventions can result in reliance on trade secret law, which reduces transparency in the field

Despite hints of forthcoming reforms to copyright in Canada, provisions to adapt IP frameworks to meet the challenges of AI have not been implemented thus far; the issues raised in earlier sections of this report, however, paint a picture of a rapidly evolving global technological landscape with laws and policies that are struggling to keep pace. At present, exceptions to copyright protection have been inconsistently proposed or enacted in some jurisdictions to promote AI development. The patentability of constituent elements of AI systems, meanwhile, will vary from one jurisdiction to another. Even within jurisdictions, ongoing court cases will impact the likelihood that an invention might be patentable. An analysis of United States Patent and Trademark Office decisions from 2013 to 2020 across different fields where AI is being applied found that the rejection rates for patent applications varied across industries and that court decisions for certain landmark cases have had dramatic ramifications for subsequent decisions (Gaudry & Vandsburger, 2020). Consequently, there is uncertainty about how effective patents and copyright are when it comes to the protection of IP in AI. This challenge will be further complicated by the eligibility issues described earlier, in situations where humans are absent from the design and discovery loop.

Concerns have been raised that the use of less formal IP protection could reduce transparency and trust in AI systems, as well as impoverish the public domain (WIPO, 2020; Hagen, 2021). In the intangible economy, trade secrets can be appealing, not only because of the uncertainty surrounding the patentability of machine-generated inventions but also because trade secret protection has become stronger in certain jurisdictions (Ciuriak & Ptashkina, 2021). Although uncertainty also exists regarding the robustness and eligibility of trade secret protection for algorithms and data (Hagen, 2021), competitive advantage can be maintained so long as these remain confidential. Confidentiality may be easily maintained, especially where AI systems are difficult to reverse engineer (WIPO, 2020).[22]

Several problematic outcomes could arise if innovators in AI adopt trade secrets as the primary means for protecting IP. First, trade secrets can result in decreased public disclosure of discoveries (Hagen, 2021), including the reduced disclosure of "negative trade secrets." Negative trade secrets refer to trade secret protection for information about inventions or approaches that do not work or are unsuccessful. This information can be as valuable as the knowledge of promising or successful approaches because it avoids dead ends in R&D (Pritt, 2018). Second, there are

22 Efforts to extract machine learning models do exist through model theft (Section 5.7), with powerful proprietary AI tools provided on a fee-for-usage basis representing potential targets (Brodt *et al.*, 2021).

concerns that incumbent players in the AI market would benefit disproportionately from confidentiality and that there may be ethical or social reasons to suspend trade secret protection so as not to exacerbate pre-existing technology gaps at the global scale (WIPO, 2020). Finally, trade secret protection can also run contrary to several emerging regulatory requirements, such as the "right to explanation" in automated decision-making (ADM) systems (Hagen, 2021).

Trade secret protection is made more fragile by the high labour mobility in the AI field and the open innovation paradigm (Ciuriak & Ptashkina, 2021). Open-source software platforms are popular in machine learning implementations. Many of these platforms — developed by industry giants and emergent from scientific research — are among the most advanced tools freely available (Engler, 2021). In this way, the true value of IP will lie in combinations of software and data. Even if both software and data are technically open, the most useful data, or the combination or means for combining software and data, can be hidden as a trade secret. Some solutions balance the desire for practitioners to maintain confidentiality while allowing for necessary levels of transparency. For example, regulators can act as third-party intermediaries when disclosure is needed for the purposes of accountability (as in the context of ADM), but secrecy is required due to information being considered a trade secret (Hagen, 2021). Policy-makers are nevertheless challenged to define incentives and accountability mechanisms to promote disclosure and avoid the problematic outcomes described above.

> "Even if both software and data are technically open, the most useful data, or the combination or means for combining software and data, can be hidden as a trade secret."

Collaborations in AI are common and involve stakeholders with diverse priorities and incentives; early and proactive agreements to clarify IP rights among collaborators can promote the commercialization and retention of IP in Canada

As noted in Sections 2.1 and 6.1, there is a great degree of interdisciplinarity in AI R&D (The Royal Society, 2017), and collaborations are frequent both in Canada and abroad (Wu *et al.*, 2020). By design, CIFAR's Pan-Canadian AI Strategy has fostered new connections in the AI research network through its funding programs, spurring increased foreign investment (Chowdhury *et al.*, 2020) and the creation of large PPPs with multinational AI leaders (Brandusescu, 2021). Although such partnerships are instrumental for creating a critical mass of AI R&D in Canada, some argue that the prominence of large foreign players in the field presents risks that could undermine "made in Canada" approaches (Brandusescu, 2021).

IP and data are highly valuable to these players, who possess a capacity to leverage these assets that is difficult to match domestically (Gallini & Hollis, 2019). Foreign ownership of IP rights developed in Canada follows frequently from PPPs, as well as through the foreign acquisition of Canadian start-ups (Hinton & Cowan, 2018). This can result in licensing challenges for SMEs operating in Canada, which may be forced to pay licensing royalties for IP-protected works developed at home or restricted from obtaining licensing at all (Hinton, 2020). Concerns surrounding the extent to which publicly funded research results in IP that ultimately leaves Canada are not limited to AI, but AI is a field where this issue is prominent. As of 2020, foreign companies had acquired the majority of AI and machine learning patents filed at the United States Patent and Trademark Office by researchers or institutions in Canada since 2015 (Hinton, 2020).[23]

Medeiros *et al.* (2021) stress the importance of contracts in overcoming issues surrounding rights and ownership of Canada-made AI advancements, but also in avoiding future conflicts among partners. Contracts — including licensing, commercial, and collaboration agreements — can play an important role in structuring relationships or governing access to and use of data or technology. For instance, collaboration agreements can be used to specify what each member contributes at the outset of a collaboration (e.g., hardware and software components). These agreements can also define the rights in subsequent IP-protectable works. This safeguard can ensure that one party in a project does not effectively cede control of technology to another, for instance, by sharing their expertise (Medeiros *et al.*, 2021). Partners might use contracts to delineate the intended use of technologies developed during the partnership (e.g., commercial or non-commercial use). In the event of previously unforeseen applications, this can be beneficial for smaller entities wishing to avoid litigation or liabilities associated with unintended uses.

Opportunities to better protect the interests and IP of non-commercial stakeholders or smaller entities also exist at the licensing stage. For example, because these actors may lack the resources to take part in extended litigation with multinational enterprises, the collaboration agreements underpinning large PPPs could be used to stipulate that SMEs based in Canada are permitted to obtain licences to IP rights generated from the larger partnership, perhaps without owing royalties for a limited duration of time (Bawa & Tawfik, 2020). This model was employed in the development of COVID-19 vaccines carried out through a PPP in the United Kingdom. Enhanced participation in patent pools — such as those created as part of the Pan-Canadian IP Strategy — and greater leveraging of IP

23 Some Canadian AI research institutions have published IP policies for potential collaborators to consider. The Vector Institute's policy describes commercializing its IP in a way that "emphasizes economic development through the creation of Canadian startups" (Vector Institute, n.d.), whereas Mila advertises a policy of public disclosure of IP through publication over patenting (Mila, 2021).

expertise found in superclusters could also facilitate more effective exploitation of IP developed in Canada through partnerships with foreign multinational enterprises (Bawa & Tawfik, 2020).

6.3 Emerging Legal Risks

The deployment of an AI system has the potential to broaden its interactions with society. Although scientists, engineers, and data scientists (among others) may spend months or years testing and training an AI system, unforeseen consequences can occur once the system is applied to a new environment. Many experts are involved in developing and deploying AI systems, such that a single "natural person"[24] deemed responsible for the system or its actions might be impossible to determine. Circumstances are also a factor; the systems developed for design and discovery in science and engineering may be specialized in scope and not as susceptible as others to problematic outcomes in highly public manners (Schwartz, 2019). However, these systems are likely to become integrated into their environments, and practitioners must consider the potential for infringement and harms, as well as the need to maintain secure AI systems given their ability to inform or carry out decision-making functions.

The opacity of some AI systems frustrates the application of legal liability principles, such that the enforcement of IP rights and privacy protection are no longer straightforward

AI systems require access to vast amounts of data and, in some cases, will rewrite segments of code in the algorithms responsible for their function. In such cases, the AI system runs the risk of infringing on existing IP rights by accessing, copying, or manipulating datasets protected by copyright or by implementing processes that are patent-protected (Medeiros *et al.*, 2021). Conventionally, IP rights-holders can sue the entities responsible for infringement to enforce their rights. However, in doing so, it is crucial to substantiate who is responsible for infringement, how responsibility is divided among multiple parties if applicable, and which actions carried out by the system resulted in the infringement (Benhamou & Ferland, 2021).

As noted above, multiple actors are typically involved in the creation of an AI system or tool, and it might be impossible to assign liability to any individual along the development life cycle (Giuffrida, 2019). Moreover, AI and its data ecosystem (as well as the devices that create and collect these data) are not separable, further complicating the assignment of liability (Giuffrida, 2019). Finally, if the AI system is

24 A "natural person" is defined as a human being, as opposed to a "legal person" or an "artificial person," which may include private or public organizations.

a black box, it may not be possible to provide the necessary cause-and-effect relationship linking specific actions by the system to the resulting infringement. Lack of transparency may prevent the identification of copyright infringement during the training phase, much less the attribution of those actions to individuals (Craig, 2021). Whether a human is in or out of the loop, and whether the AI is considered a product or a service (or embedded in a device), will dictate how liability might foreseeably be assigned (Giuffrida, 2019).

> "The question is whether existing legal frameworks should adapt or whether new legal frameworks are required to regulate AI."

Though some research activities might be exempted from the risk of litigation,[25] there remain concerns about the potential for unpredictability in IP enforcement (Medeiros *et al.*, 2021). In the absence of new legislative measures such as TDM exemptions in the *Copyright Act* (Section 6.1), Medeiros *et al.* (2021) argue that contractual agreements can help to fill current gaps. Contracts and licensing agreements between legal entities can define rights and obligations surrounding the use of copyright-protected works by AI systems. Licensees and users, meanwhile, could also benefit by conceivably being exempt from certain IP infringements, leaving the product owners to address claims of that nature (Medeiros *et al.*, 2021). Although contracts do not resolve broader issues surrounding liability in IP infringement, IP holders may find this tool helpful as a mechanism to address protection and accountability.

Legislators and legal scholars are considering frameworks for liability and accountability that balance the desire to assist in the development of AI systems and the need to protect against harm

AI systems can conceivably cause significant harm that may result in legal action. For example, AI systems that monitor or maintain infrastructure (Sparkes, 2021), and those that create new forms of artificial life that might integrate with a natural ecosystem (Kriegman *et al.*, 2020) can carry risks of potential harm resulting from applying AI in the real world. Legal regulation will be expected to protect people from such harms while also encouraging innovation. As tort liability determined by judges applying broad legal principles to new scenarios introduced by AI systems may not be the most efficient or effective means of compensating for harm caused by AI, legislative reform is inevitable (Giuffrida, 2019). The question is whether existing legal frameworks should adapt or whether new legal frameworks are required to regulate AI (Giuffrida, 2019).

25 For instance, activities that fall under fair-dealing limitations for copyright which allow for the "fair" use of copyright-protected material (Craig, 2021).

Liability principles differ depending on whether they are applied to products or to services. Determining which liability regime would apply to an AI system depends on several factors, including whether the AI constitutes a "tangible form" (Chagal-Feferkorn, 2019). Although the concept of "autonomy" is widely used to distinguish traditional from sophisticated technologies, it is not necessarily useful for determining which liability regime should apply (Chagal-Feferkorn, 2019). Product liability may not adequately govern AI for several reasons. AI can cause harm in the absence of any defect in its system, a necessary element of product liability (Benhamou & Ferland, 2021). Furthermore, the transparency problem associated with black box AI systems limits the foreseeability of its outputs, another necessary element of product liability (Benhamou & Ferland, 2021).

Several solutions have been proposed in response to the difficulties in reconciling conventional liability frameworks with processes and outputs accompanying AI systems that possess varying levels of autonomy (Benhamou & Ferland, 2021). Existing liability regimes for products and services could be extended to automatically hold the producers or operators of the AI systems liable through various mechanisms (Benhamou & Ferland, 2021). Alternatively, granting legal personhood status to AI would resolve some of the identified issues, but it may also have important ramifications for copyright and patent law (Section 6.2). Granting legal personhood status to AI would also, however, raise its own challenges, given the previously mentioned difficulties in identifying a "natural person" responsible for the AI (Benhamou & Ferland, 2021).

There is presently a lack of consensus internationally with respect to the path forward, but the proposed EU regulation for AI provides insight on how that jurisdiction intends to proceed. Legal personhood is not currently viewed as a viable solution, nor is an overhaul of the existing liability regime in the EU and its member states (European Parliament, 2020). Rather, the European Parliament sees the revision of product liability frameworks as a path forward, as well as "strict liability" for autonomous AI systems that are deemed to be high risk (European Parliament, 2020; European Commission, 2021a). In this framework, operators of AI systems would be held liable for harm, even if they are not demonstrably at fault, due to the inherent risks arising from the uncontrollability and unpredictability of the system they are using (European Parliament, 2020). Giuffrida (2019) argues that risk-based or harms-based approaches for tackling liability have parallels with existing approaches for regulating privacy and cybersecurity and would offer a straightforward system for ensuring compensation for harms through some form of liability insurance (European Parliament, 2020). The approach pursued by the EU converges on a harmonized approach and recognizes the risk inherent in a regulatory vacuum surrounding liability from the standpoints of human rights and economic development (European Parliament, 2020).

6.4 Emerging Regulatory Systems for AI

As discussed, there are perceived tensions between regulating AI and fostering an innovation-friendly environment for AI in other industries. There are numerous stakeholders in this context, each with a distinct role: policy-makers can provide directions as to what they consider *should* be done to regulate AI, but what *can* be done will depend strongly on the state of the law. The interplay among policy-makers and legislators and those of outside developers and users of AI present numerous friction points. Thus far, regulatory interventions in the case of AI — particularly proactive ones — are seen as supporting innovation by positively contributing to trust in AI by multiple stakeholders (Deloitte Canada, 2019; LCO, 2021). These interventions can also mitigate legal uncertainty in the deployment of AI systems that might otherwise dissuade investment (Giuffrida, 2019). For example, the EU offers a large "digital single market" through numerous laws and policies, most importantly through the GDPR but also through its recent proposal for a comprehensive regulatory framework for AI (Mercurio & Yu, 2021).

Several Canadian approaches for the responsible development and deployment of AI and its associated technologies now exist in a complex but incomplete landscape, and others are in development (LCO, 2021). This complexity is due in part to the division of powers in Canada's federal constitution and results in various aspects of AI regulation being distributed across federal and provincial/territorial jurisdictions. Provincial and territorial governments possess considerable autonomy in several areas relevant to the development of AI, such as healthcare and education. Jurisdiction over privacy and data protection is also fragmented across public, private, and health sectors. A regulatory environment for AI in Canada is beginning to take shape through several policy developments occurring asynchronously at different levels of government. Given the international dimension of the governance and development of AI (Sections 6.1 and 6.2), local decisions made in Canada can have global implications and vice versa. Values and norms will be reflected in regulatory environments, with implications for collaboration and trade and the potential for incompatibilities across borders.

Although the Government of Canada has implemented a framework for AI governance, its scope is limited to the federal public service, resulting in a regulatory gap

One framework for governing AI in Canada operates at the federal level as part of the Responsible Use of AI initiative (GC, 2021f). It consists of a set of guidelines, a list of qualified suppliers eligible to provide services to government departments and agencies using AI systems, an algorithmic impact assessment tool, and the

Directive on Automated Decision-Making (DADM) (GC, 2021f). The DADM is a lynchpin of the framework due to the capacity for AI systems to negatively impact human rights and erode trust, and it has been lauded for its proactive nature (LCO, 2021). However, its scope of application is inherently limited: it does not extend to provincial, territorial, or municipal governments due to the division of powers in Canada. It also does not apply to private sector activities unless they relate to products or services procured by the federal government (LCO, 2021). This results in a regulatory gap in Canada (LCO, 2021). Despite ongoing initiatives by some provincial/territorial governments to develop equivalent frameworks (Gov. of ON, 2021c), many potential applications of ADM (and other uses of AI) by provincial, territorial, and municipal agencies are currently unregulated.[26] The current regulatory environment for AI in Canada has been criticized by some for having an "ad-hoc approach" and for lacking coherence across the numerous relevant orders of Canadian government (Brandusescu, 2021; McKelvey & Roberge, 2021).

The laws governing access to certain types of personal data in Canada contribute to the regulation of AI as well, albeit indirectly, in an area that is also complicated by the division of powers. Privacy laws intended to govern the use of personal data are fragmented according to the type of organization (public or private sector) making use of such data, where the data are located, the type of data, and whether the data cross provincial/territorial borders (OPC, 2021a). The lack of harmonization contributes to challenges for data protection and data sharing. Attempts to reform and modernize legislation are underway to account for the opportunities and risks presented by AI and big data, particularly regarding commercial activities. At the federal level, Bill C-11 (the *Digital Charter Implementation Act*) sought to revise federal privacy legislation applicable to personal data used by the private sector (Cofone *et al.*, 2021), but did not proceed as a federal election was called in the autumn of 2021 (Smith, 2021). In Quebec, meanwhile, the recently enacted Bill 64 (Assemblée Nationale, 2020) aims to address shortcomings in previous legislation that was deemed "ill-suited to the current context resulting from large-scale development and adoption of digital technologies" [free translation] (CEST, 2021). Areas where the previous legislation was seen as inadequate included the handling of personal information that might be inferred by AI systems, the means by which data could be used for research, and the lack of provision for individual rights relating to automated decision-making. It should be noted that Ontario has proposed enacting its own private sector data protection law, which would draw upon provisions in both Bill C-11 and Quebec's Bill 64 (Gov. of ON, 2021a).

26 Governments in the United States, the United Kingdom, and New Zealand currently use ADM tools to assist numerous operations in the areas of criminal, administrative, and civil justice (LCO, 2021).

Applications of AI systems in science and engineering for design and discovery might not consistently make use of personal data and therefore may be outside of the focus of the above reforms. Nevertheless, the Panel believes that the proposed reforms to laws governing private sector data could act as a nucleus for a broader regulatory framework moving forward.

The challenges in regulating AI are prompting the use of novel regulatory approaches that will be subject to ongoing review and renewal

There are numerous scenarios where legal or regulatory tools prove inadequate to address situations raised by the deployment of AI systems outside of a laboratory research setting. Reforms or adjustments to existing laws and processes are ongoing in multiple jurisdictions to prepare for increasingly prevalent and autonomous AI systems, but some are also exploring new regulatory tools that might be more adaptable or responsive on account of the rapid evolution of technology. Several examples of experimentation are occurring in the form of regulatory sandboxes, which are designed to facilitate real-world testing for innovators in a controlled way, such that the learnings can provide regulators with input for policies. Sandboxes take on numerous forms, from time-limited regulatory waivers to specific physical locations where experimentation can take place (OECD, 2021b). The latter examples vary in scale and have been established in research institutions and laboratories, along roadways, or across entire geographic regions.[27] Several countries have either developed sandboxes or intend to do so based on their national strategies (Kung *et al.*, 2020). A regulatory sandbox is already being introduced in Canadian healthcare for therapeutic products based on new technologies, including those using AI (HC, 2019). The details of the implementation are not fully known (HC, 2021), but some features of the new mechanism have already been applied in response to the COVID-19 pandemic through clinical trial design and reporting requirements (Eren Vural *et al.*, 2021). A key aspect relates to how much evidence needs to be provided to regulators, and how early, before a product can start being used by patients (HC, 2019). Similar strategies could be applied in regulatory frameworks for AI systems to potentially allow quicker entry to market; this would, however, place pressure on governments to manage risks and scrutinize post-market outcomes to inform the necessary adjustments to regulations.

Regular reviews and renewal of existing AI policy can help to avoid negative outcomes, as well as take advantage of new opportunities. The OECD's AI Principle 2.3 recommends that the deployment of trustworthy AI systems should include the identification of mechanisms for "improving the adaptability, reactivity,

27 China has deployed a "Pilot Zone" for AI development across a county with 500,000 inhabitants, where flexible regulations will allow for experimentation relating to autonomous vehicles, smart agriculture, and AI in government (Xinhuanet, 2020).

versatility and enforcement of policy instruments" that apply to AI (OECD, 2021d). The prevalence of self-regulatory approaches thus far only heightens the need for transparency between developers and regulatory authorities so that reviews and evaluations can take place (Renda, 2019). In Canada, the DADM currently follows a six-month review cycle conducted through a peer review process involving stakeholders from both inside and outside of the public sector (GC, 2021f). The European Commission encourages member states to review and update national AI strategies as necessary (European Commission, 2021b). Several EU countries — including Cyprus, Denmark, Italy, Latvia, Luxembourg, Malta, the Netherlands, Poland, Portugal, Slovakia, and Sweden — have explicitly indicated that their AI policies and strategies will be regularly reviewed and updated, typically on an annual basis. In addition, France has indicated its intention to set up a national platform for auditing AI algorithms (Van Roy, 2019). The European Commission also requires that providers of "high-risk AI systems"[28] develop processes and indicators to monitor the design, development, and testing of these systems, which must be periodically audited by a "notified body" as a form of quality control (European Commission, 2021b).

It will also be important for governments to periodically review and modify their policies in consideration of their practical impacts and new developments in the field of AI. This will also be true for other important stakeholders such as universities, funding agencies, and industry. Doing so will require decision-makers to be mindful of the variable timescales of their policies with respect to impact. The consequences of policies surrounding research funding and innovation may appear more slowly than those relating to procurement or governing the use of AI in intramural research. Multisectoral participation in reviews may also add value as a means of diffusing AI by raising awareness of challenges and opportunities and avoiding the appearance that regulations are developed by a shallow pool of stakeholders (Section 2.3).

Decisions made outside of Canada will create pressure to harmonize Canadian approaches with those implemented abroad

Despite its historical leadership in some AI research areas, Canada has struggled to develop its regulatory framework in isolation from its largest neighbours and trading partners. In this respect, though the DADM represented an important proactive step towards defining norms and practices at the scale of the public service, regulatory frameworks implemented in other jurisdictions are poised to impact numerous stakeholders in Canada. Thus far, approaches to regulating AI in other leading nations in AI R&D are diverse, ranging from market-led approaches

28 High-risk systems in the EU regulations describe systems that "pose significant risks to the health and safety or fundamental rights of persons" (European Commission, 2021b).

relying on self-regulation by developers, to approaches where government exerts a more prescriptive function. Self-regulation proceeds according to guidelines or standards that might be defined by the industry itself or national advisory groups (Geist, 2021) (Section 2.3). Several countries, including Canada, Israel, the United Kingdom, and the United States, are adopting a market-led approach — thus far to varying degrees — while China maintains a government-led model (Geist, 2021).

> “The diverse priorities at federal and provincial/territorial levels may lead to a flexible but ultimately fragmented regulatory environment for AI.”

The United States has most strongly embraced the market-led approach, proceeding according to soft law and self-regulation described by guidelines and codes of conduct for individual stakeholders (Castets-Renard, 2021). Examples of laws to regulate AI systems are absent at the federal level, [29] and tend to apply to specific individual sectors, only being found at the state and municipal levels, such as in transportation and law enforcement, respectively (Castets-Renard, 2021). China’s model, despite being government led, is highly decentralized in practice due to the number of layers of government involved (Roberts *et al.*, 2021). Its framework places greater emphasis on innovation, economic development, and societal benefit, with less of a focus on individual rights or safety (Roberts *et al.*, 2021).

In 2021, the EU unveiled its proposed regulation for AI, which lies somewhere between the U.S. and China on the regulatory spectrum (Geist, 2021). The regulation is designed to be technologically neutral and applies to public and private companies, as well as both ADM systems and content generated by AI (European Commission, 2021b). The regulation is risk-based, similar to the GDPR, where risks are defined with respect to safety, human rights, uncertainty, and specificity (European Commission, 2021a). Risks are assigned a level of severity, resulting in proportionally increasing requirements placed on developers and vendors. Risk-mitigation strategies are based on data governance and documentation, and the EU foresees the use of regulatory sandboxes to provide SMEs with an opportunity to test the new framework in the context of low-risk AI systems (Marcia & DeSouza, 2021). High-risk applications involving systems possessing high levels of autonomy will require rigorous oversight mechanisms (e.g., the proposed “strict liability” regime discussed in Section 6.3). Floridi (2021)

29 Federal initiatives have been launched in the United States, such as the National AI Initiative Office (Niczepyr, 2021), but the accompanying legislation has thus far focused on guidance for the development of AI and not regulatory instruments (NAIIO, 2021).

cautions, however, that the regulation does not sufficiently distinguish between risk resulting from failure or error and risk resulting from the intent of the user.

By being the first to disclose a proposed framework, the EU hopes to assert leadership in developing AI responsibly (European Commission, 2021a). Nevertheless, there are also risks associated with taking a strong regulatory stance, given the lack of international harmonization. For example, the EU might inadvertently direct the development of ethically problematic AI applications elsewhere, but these applications could subsequently be re-imported after being made domestically compliant (Floridi, 2021). This evolving international regulatory environment impacts Canadian activity in AI for science and engineering, given the divergence in approaches among some of Canada's largest partners in research and trading alike (Geist, 2021). This might result in pressure to align Canada's approach with that of another jurisdiction in order to minimize inconsistencies. In response, federal and provincial/territorial governments may try emulating regulations proposed by the EU, emphasizing the importance of rights and prevention of societal harm (Castets-Renard, 2021). Alternatively, the diverse priorities at federal and provincial/territorial levels may lead to a flexible but ultimately fragmented regulatory environment for AI. Dawson (2021) suggests that Canada could yet lead a multilateral initiative aimed at harmonizing regulations and standards, but until such an initiative occurs, Canadian AI researchers will need to be mindful of how divergent international norms impact their collaborations, particularly with respect to data governance. Similarly, Canadian SMEs (which may feel pressure to access certain markets), will need to balance compliance with international regulations and domestic uncertainty.

7

Conclusion

Realizing the promise of integrating AI with science and engineering will allow for new questions to be examined and will accelerate innovation in numerous other technological areas. This report explores the broad spectrum of legal/regulatory, ethical, social, and policy issues in the context of the design and deployment of AI for science and engineering.

7.1 Addressing the Charge

With continuing algorithmic advances and the wide availability of computational resources and scientific data, ideal conditions are emerging for the application of AI to design and discovery in science and engineering. However, despite an increasing number of reported breakthroughs demonstrating the promise of AI, several real and imminent challenges must first be overcome. To help understand these opportunities, challenges, and implications, the NRC, with support from CIFAR, CIHR, NSERC, and SSHRC, asked the CCA to convene a multidisciplinary and multisectoral expert panel to answer the following question:

What are the legal/regulatory, ethical, social, and policy challenges associated with deploying artificial intelligence technologies to enable scientific/engineering research design and discovery in Canada?

In this section, the Panel addresses this question and highlights its overarching conclusions.

The increased use of AI will shift epistemic, ethical, and institutional practices in science and engineering

Addressing challenges related to the accuracy, explainability, and reproducibility of the results generated by AI systems will require increased transparency on the part of researchers and the establishment of new data management standards and practices to protect scientific integrity. Furthermore, even if the results themselves are reliable, the complexity and opacity of certain types of AI systems raise questions about whether users will be able to understand or explain any innovative findings or novel designs that they might obtain by using these systems.

Ethical questions about the use of AI for science and engineering arise at all stages in the process, including data collection and pre-processing; the design and deployment of AI models trained on those data; the dissemination and publication of results; and the long-term storage, maintenance, and access to data, models, and results. The importance of data, especially big data, presents challenges for many aspects of traditional research ethics, particularly for research involving human participants and historically marginalized groups. To address these issues, institutional actors (including funding agencies, universities, and R&D firms) may need to update their frameworks for conducting ethical research to account for the implications of AI. For example, the use of AI in science and engineering may impact social groups more than individuals, requiring a shift in focus away from individual harms and towards social harms. However, traditional research ethics review boards are ill-equipped to consider such impacts.

> "Even if the results themselves are reliable, the complexity and opacity of certain types of AI systems raise questions about whether users will be able to understand or explain any innovative findings or novel designs that they might obtain by using these systems."

The increased use of AI in the fields of science and engineering could also change their social dynamics. For example, the social practices underlying the dissemination of scientific findings (publications, conferences), establishing their validity (peer review, replication), and acknowledging their provenance (crediting, citations) may all need to be revised due to the increased use of AI in scientific research. The increased use of AI will transform the nature of scientific inquiry and thus require human scientists and engineers to work differently. By allowing machines to guide the directions of future research, some limitations of human decision-making might be overcome, but at the expense of control.

Approaches emphasizing equitable access, diversity, and inclusion will help address the social and ethical challenges associated with the use of AI in the Canadian research system

There are significant inequities in both the field of AI research itself and in access to AI technology. The current lack of gender and racial diversity in the field of AI is well documented, and there are high levels of inequality in the existing distribution of resources, infrastructure, and skills in the context of AI for scientific research. Ensuring equitable access to AI is favourable from multiple standpoints and is a critical issue for both governments and research institutions to address. On the one hand, improving access to the resources needed to develop or apply AI systems to science and engineering will facilitate the diffusion of the technology across domains, with the promise of enabling a wider range of applications. On the other hand, access is also a key factor from the standpoint of the distribution of benefits relating to AI systems already being deployed.

In both cases, strong incentives for inclusion and transparency would improve equity in AI research and access in Canada. The exacerbation of existing inequalities in access to AI resources may be challenging to avoid, given the level of influence early adopters can exert over its rapid development. Gaps in the Canadian research and innovation environment risk widening unless care is taken to promote access to tools, resources, and employment in this emerging sector.

Research funding agencies may need to adapt certain evaluation practices because the allocation of public research funds will dictate which types of AI applications in science and engineering will be prioritized and what standards of conduct will be expected from funding recipients. The research funding system may additionally be tasked with integrating AI tools into its processes for assessing and evaluating research and research impacts. It will also be asked to implement programs and policies to cultivate new interdisciplinary partnerships and promote the creation and sharing of high-quality data. It will be particularly important, therefore, for funders to monitor issues surrounding bias and discrimination when adopting AI in processes that manage research funding. The credibility of such initiatives will hinge on the establishment of similar practices for R&D in order to ensure accuracy, interpret uncertainty, and promote trust.

> “The automation of science and engineering resulting from the integration of AI could be seen as a threat to the research labour market in Canada because at least some scientific and engineering occupations are exposed to potential displacement by AI.”

As it proceeds through an uncertain legal and regulatory environment, the use of AI in science and engineering will raise questions about social benefits and harms

The increased use of AI tools in science and engineering is likely to have significant implications for broader society. Although discoveries made using AI systems could help to address the climate crisis, the significant environmental impacts of the development and operation of AI remain an unresolved concern. Furthermore, the automation of science and engineering resulting from the integration of AI could be seen as a threat to the research labour market in Canada because at least some scientific and engineering occupations are exposed to potential displacement by AI. AI systems have also acted as amplifiers for existing social challenges and inequities in many areas where they have been deployed thus far. The repeated manifestation of discriminatory outcomes against historically marginalized groups is a well-known and dangerous flaw exhibited by AI systems and could be perpetuated by their use in science and engineering. Indigenous communities may be asked to participate in scientific collaborations using AI systems that were designed without their input, with risks that culturally appropriate practices with respect to research conduct and data are not followed.

Some of these risks emerge because there are gaps between the principles for the responsible development of AI and their operationalization, as well as a paucity of stronger regulatory measures overall. In Canada and internationally, the legal and regulatory environment is currently struggling to keep pace with technological progress, and the desire to facilitate innovation exists in tension with the duty to protect society from harm. This tension manifests itself in numerous ways, ranging from compromises between protecting and sharing sensitive data to decisions about the attribution of liability for autonomous systems. This environment also reflects struggles to reach consensus on the legal status of discoveries and designs produced by AI systems, with implications for commercialization and access. In areas where regulations are currently absent, contracts and licences may help to delineate rights and responsibilities for innovators. Efforts to harmonize and modernize legal and regulatory frameworks at national and international scales may face hurdles due to diverse cultural and political values and path-dependency, leading to a heterogenous landscape that will influence how and where applications are developed.

> "The legal and regulatory environment is currently struggling to keep pace with technological progress, and the desire to facilitate innovation exists in tension with the duty to protect society from harm."

National AI policies and strategies that cut across multiple policy domains may bolster developments in science and engineering

A growing number of countries, including Canada, have developed national strategies for AI. The Canadian strategy has traditionally focused on AI research and capacity building but has more recently begun to target multiple sectors and consider societal impacts more explicitly. Broadening the scope of AI strategies can lead to new connections within the R&D network, and help to establish crucial links between AI policy and areas outside of the AI environment. For example, policies promoting access to data and open data are not strictly AI policies, but are important enabling elements. Investments in data infrastructure and data policy initiatives are correspondingly multiplying either as explicit components of recent international AI strategies or as complementary efforts.

Decisions and developments in numerous policy dimensions will have an impact on both the speed and trajectory of growth in the use of AI for science and engineering. Indeed, some scientific research using AI already dovetails with existing policy goals and initiatives — efforts to combat climate change through innovation, for instance, are spurring funding towards materials science research using AI. Other policy trends, such as shifting approaches in the assessment of research and researchers, are adjacent to AI, but will also influence its development. To move AI in science and engineering forward, it will be helpful to proactively identify and manage the interconnections among policy areas to account for the wide variety of stakeholders beyond the AI community.

7.2 Panel Reflections

Research and commentary on the responsible development and use of AI have typically focused on applications with the potential to infringe on human rights or privacy. By contrast, the development and use of AI for science and engineering research have been comparatively less discussed, although looming concerns in these disciplines mirror many of those already being considered for society at large. Moreover, although innovation in science and engineering is at times seen as an intrinsic good, developments in AI technology for these disciplines would also lead to concomitant advances of AI in other more controversial areas. Establishing good practices and addressing the amplification of social and ethical issues by AI systems used in science and engineering may provide lessons on avoiding unwanted outcomes in other areas of society where AI is being deployed.

> "Training future scholars and engineers to appreciate ethical and social dilemmas could prevent potentially unethical or unintended outcomes, and subsequently reinforce public trust in discoveries made using AI."

The cross-cutting nature of AI highlights the need to develop expertise and roles to better integrate knowledge and skills across multiple traditional disciplines. This expertise in managing data, or in reviewing AI tools, together with policies to incentivize transparency, will help address several epistemic and ethical challenges that exist at the conceptual and developmental stages of AI tools for science and engineering. Similarly, the research funding system can better support AI R&D in Canada by adapting and learning to assess and evaluate research and researchers through different lenses. For example, the concept of *research excellence* may need to be revised if AI systems take on greater roles in driving research. Training future scholars and engineers to appreciate ethical and social dilemmas could prevent potentially unethical or unintended outcomes, and subsequently reinforce public trust in discoveries made using AI or in objects, materials, and processes designed by AI. Deployment, meanwhile, will continue to be complicated by lingering uncertainties surrounding several unresolved legal and regulatory issues. Decision-makers would benefit from broader consultation across jurisdictional and professional lines as they wrestle with choices surrounding data governance, acceptable levels of societal risk, and how to manage IP rights. Although reforms in the latter area are ongoing, regular reviews and renewal of existing AI policies in all areas will help to avoid negative outcomes (or learn from them), as well as take advantage of new opportunities.

“This fluid situation offers opportunities to look to the future and pinpoint areas along the life cycle of AI development and deployment where decision-makers might intervene before predicted hurdles gradually materialize.”

This report identifies the large number of stakeholders whose actions and decisions will determine how the above challenges are addressed and who will shape how disparate fields and sectors might integrate AI into their practices. This fluid situation offers opportunities to look to the future and pinpoint areas along the life cycle of AI development and deployment where decision-makers might intervene before predicted hurdles gradually materialize. At the front end, research funding agencies, academic institutions, and public agencies tasked with supporting innovation can take proactive steps to lower barriers for interdisciplinary and intersectoral collaboration between AI researchers/practitioners in science and engineering and those working in other disciplines. The tailoring of funding practices, education and training, and access to AI resources will each contribute to Canada’s capacity to leverage strengths in AI research developed over the previous decade(s) towards the use of AI for design and discovery in science and engineering. Failure to adapt existing practices in these and other areas will complicate any efforts to broaden Canada’s AI ecosystem beyond the current major centres where R&D is most active, with the potential for missed opportunities to grow horizontally outside of fundamental AI research. After all, the present networks comprising the Canadian R&D landscape will not necessarily evolve to make the most effective use of AI for science and engineering without new and possibly unexpected connections being made. The extent to which Canadian society may benefit from the decades of public investments in AI will hinge on these connections and the resulting uptake of AI tools throughout the innovation community.

"The extent to which Canadian society may benefit from the decades of public investments in AI will hinge on these connections and the resulting uptake of AI tools throughout the innovation community."

Importantly, however, realizing the benefits of employing AI for science and engineering research will depend on ensuring that these systems are used carefully, responsibly, and wisely. Whereas AI may offer the potential to transcend the limitations of human cognitive abilities by producing novel scientific discoveries and innovative engineering designs, it also has the potential to perpetuate human biases, as well as create entirely new ones. If not used responsibly, AI could exacerbate inequalities in both the research and innovation systems as well as in broader society. To avoid such negative outcomes, social and ethical considerations will need to be addressed, not just at the deployment stage but in the very earliest stages of AI development. Moreover, the responsible use of AI for science and engineering will also require transparency and oversight mechanisms to ensure that results generated by AI are accurate, reproducible, and explainable in ways that further scientific understanding.

References

Acceleration Consortium. (2021). Mission. Retrieved June 2021, from https://acceleration.utoronto.ca/#mission.

Adadi, A. & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138-52160.

AI HLEG (High-Level Expert Group on Artificial Intelligence). (2019). *Ethics Guidelines on Trustworthy AI*. Brussels, Belgium: European Commission.

Amit, R. & Meir, R. (2019). Lifelong learning and inductive bias. *Current Opinion in Behavioral Sciences*, *29*, 51-54.

Amnesty International & Access Now. (2018). *The Toronto Declaration - Protection the Right to Equality and Non-Discrimination in Machine Learning Systems*. Toronto (ON): Amnesty International and Access Now.

Ananthaswamy, A. (2021a, April 19). Latest Neural Nets Solves Worlds Hardest Equations Faster than Ever Before, *Quanta Magazine*.

Ananthaswamy, A. (2021b, July 2). AI Designs Quantum Physics Experiments Beyond What Any Human Has Conceived, *Scientific American*.

Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker - Tracking and Predicting the Carbon Footprint of Training Deep Learning Models [pre-print]. *arXiv*(2007.03051v1).

Anyaha, R. (2017). The History of Artificial Intelligence. Retrieved May 2021, from https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

ARDC (Australian Research Data Commons). (2021). Research Data Australia. Retrieved July 2021, from https://researchdata.edu.au/page/about.

Asay, C. D. (2020). Independent Creation in a World of AI. *Florida International University Law Review*, *14*(2).

Assemblée Nationale. (Assemblée Nationale du Québec). (2020). *Projet de loi no 64 — Loi Modernisant des Dispositions Législatives en Matière de Protection des Renseignements Personnels*. Québec (QC): Government of Québec.

Australian Government. (2021). National Data Commissioner. Retrieved August 2021, from https://pmc.gov.au/public-data/national-data-commissioner.

Baker, D., Bourne-Tyson, D., Gerlitz, L., Haigh, S., Khair, S., Leggott, M., . . . Whitehead, M. (2019). *Research Data Management in Canada: A Backgrounder.* Canadian Association for Research Libraries (CARL), Consortia Advancing Standards in Research Administration Information (CASRAI), Leadership Council for Digital Research Infrastructure (LCDRI), Research Data Canada (RDC).

Bates, J. (2021). Expanding the Geography of Innovation: NSF AI Institutes 2021. Retrieved August 2021, from https://beta.nsf.gov/science-matters/expanding-geography-innovation-nsf-ai-research-institutes-2021.

Bawa, K. & Tawfik, M. (2020, December 30). How to Keep IP Canadian: Learn from the Success of Public-private Partnerships During the Pandemic, *The Globe and Mail.*

BCOHRC (British Columbia's Office of the Human Rights Commissioner). (2020). *Disaggregated Demographic Data Collection in British Columbia: The Grandmother Perspective.* Vancouver (BC): BCOHRC.

Bejani, M. M. & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review.*

Benhamou, Y. & Ferland, J. (2021). Artificial Intelligence and Damages: Assessing Liability and Calculating Damages. In G. D'Agostino, A. Gaon & C. Piovesan (Eds.), *Leading Legal Disruption: Artificial Intelligence and a Tool Kit for Lawyers and the Law.* Toronto (ON): Thomson Reuters Canada.

Berendt, B. & Preibusch, S. (2017). Toward accountable discrimination-aware data mining: The importance of keeping the human in the loop-and under the looking glass. *Big Data, 5*(2), 135-152.

Berman, F. (2019). *The Research Data Alliance — The First Five Years*. Research Data Alliance.

Bernstein, M. S., Levi, M., Magnus, D., Rajala, B., Satz, D., & Waeiss, C. (2021). ESR: Ethics and Society Review of artificial intelligence research. *arXiv*(2106.11521v2).

Bezuidenhout, L. & Ratti, E. (2020). What does it mean to embed ethics? *AI & Society*(36), 939-953.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns, 2*(2), 100205.

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics, 26*(7), 931-936.

Bogue, R. (2012). Robots in the laboratory: A review of applications. *Industrial Robot, 39*(2), 113-119.

Bond, T. (2021). News Flash: UK Court of Appeal Rejects DABUS Patent Application. Retrieved October 2021, from https://www.twobirds.com/en/news/articles/2021/uk/uk-court-of-appeal-rejects-dabus-patent-application.

Borrows, J. (2010). *Canada's Indigenous Constitution.* Toronto (ON): University of Toronto Press.

Borrows, J. (2012). (Ab)Originalism and Canada's constitution. *The Supreme Court Law Review: Osgoode's Annual Constitutional Cases Conference, 58*, 351-398.

Boschert, S., Heinrich, C., & Rosen, R. (2018). *Next Generation Digital Twin.* Paper presented at Twelfth International Symposium on Tools and Methods of Competitive Engineering, Las Palmas de Gran Canaria, Spain.

Boulanin, V., Brockmann, K., & Richards, L. (2020). *Responsible Artificial Intelligence Research and Innovation for International Peace and Security.* Solna, Sweden: Stockholm International Peace Research Institute.

Boutilier, A. (2021, June 10). RCMP Broke Privacy Laws in Using Controversial Clearview AI Facial Recognition Tools, Watchdog Says, *Toronto Star.*

boyd, d. & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662-679.

Brandusescu, A. (2021). *Artificial Intelligence Policy and Funding in Canada: Public Investments, Private Interests*. Montréal (QC): Centre for Interdisciplinary Research on Montreal.

Brandusescu, A., Cutean, A., Dawson, P., Davidson, R., Matthews, M., & O'Neill, K. (2021). *Maximizing Strength and Spearheading Opportunity: An Industrial Strategy for Canadian AI.* Ottawa (ON): Information and Communications Technology Council.

Brenner, S. (1995). Loose ends. *Current Biology, 5*, 568.

Brodt, O., Khavkin, M., Rokach, L., Shabtai, A., & Elovici, Y. (2021). Artificial Intelligence and (the Lack of) Security. In G. D'Agostino, A. Gaon & C. Piovesan (Eds.), *Leading Legal Disruption: Artificial Intelligence and a Tool Kit for Lawyers and the Law.* Toronto (ON): Thomson Reuters Canada.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? *AEA Papers and Proceedings, 108*, 43-47.

Caroll, M. W. (2019). Copyright and the progress of science: Why text and data mining is lawful. *UC Davis Law Review, 53*, 893–963.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., . . . Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal, 19*(43), 1-12.

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data, 8*(1), 108.

Carter, L., Liu, D., & Cantrell, C. (2020). Exploring the intersection of the digital divide and artificial intelligence: A hermeneutic literature review. *AIS Transactions on Human-Computer Interaction, 12*(4), 253-275.

Carter, R. E., Attia, Z. I., Lopez-Jimenez, F., & Friedman, P. A. (2019). Pragmatic considerations for fostering reproducible research in artificial intelligence. *Nature Partner Journals Digital Medicine,* 2, 42.

Castelvecchi, D. (2021). AI maths whiz creates tough new problems for humans to solve. *Nature, 590*, 196.

Castets-Renard, C. (2021). AI and the Law in the European Union and the United States. In F. Martin-Bariteau & T. Scassa (Eds.), *Artificial Intelligence and the Law in Canada.* Toronto (ON): LexisNexis.

CCA (Council of Canadian Academies). (2010). *Honesty, Accountability and Trust: Fostering Research Integrity in Canada*. Ottawa (ON): The Expert Panel on Research Integrity, CCA.

CCA (Council of Canadian Academies). (2012). *Strengthening Canada's Research Capacity: The Gender Dimension*. Ottawa (ON): The Expert Panel on Women in University Research, CCA.

CCA (Council of Canadian Academies). (2013). *Paradox Lost*. Ottawa (ON): CCA.

CCA (CCA). (2018). *Competing in a Global Innovation Economy: The Current State of R&D in Canada*. Ottawa (ON): The Expert Panel on the State of Science and Technology and Industrial Research and Development in Canada, CCA.

CCA (Council of Canadian Academies). (2021a). *Powering Discovery*. Ottawa (ON): The Expert Panel on International Practices for Funding Natural Sciences and Engineering Research, CCA.

CCA (Council of Canadian Academies). (2021b). *Waiting to Connect*. Ottawa (ON): The Expert Panel on High-Throughput Networks for Rural and Remote Communities in Canada, CCA.

Ceccaroni, L., Bibby, J., Roger, E., Flemons, P., Michael, K., Fagan, L., & Oliver, J. L. (2019). Opportunities and risks for citizen science in the age of artificial intelligence. *Citizen Science: Theory and Practice, 4*(1), 29.

CEST (Commission de l'Éthique en Science et en Technologie). (2021). *Mémoire de la Commission de l'Éthique en Science et en Technologie — Consultations particulières sur le projet de loi no 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*. Québec (QC): Government of Québec.

Chagal-Feferkorn, K. A. (2019). Am I an algorithm or a product? When products liability should apply to algorithmic decision-makers. *Stanford Law & Policy Review, 30*, 61–113.

Chang, T. Y.-C. (2021). DABUS the AI is Denied Inventorship in Taiwan. Retrieved October 2021, from https://www.mondaq.com/patent/1108642/dabus-the-ai-is-denied-inventorship-in-taiwan.

Charlin, L. & Zemel, R. S. (2013). *The Toronto Paper Matching System: An Automated Paper-Reviewer Assignment System.* Paper presented at The 30th International Conference on Machine Learning, Atlanta (GA).

Chawla, D. S. (2020, November 23). Do You Have A Conflict Of Interest? This Robotic Assistant May Find It First, *The New York Times*.

Chawla, D. S. (2021, May 21). Frosty Reception For Algorithm That Predicts Research Papers' Impact, *Nature*.

Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications, 8*(1), 25.

Chen, M., Tworek, J., Jun, H., Yuan, Q., Ponde de Olieira Pinto, H., Kaplan, J., . . . Zaremba, W. (2021). Evaluating large language models trained on code. *arXiv*(2107.03374v2).

Chowdhury, N., Hakim, Z., Taylor, N. A., Remennik, T., Rogers, S., Strome, E., & Wallis, J. (2020). *Pan-Canadian AI Strategy Impact Assessment Report*. Toronto (ON): Accenture & CIFAR.

CIFAR (Canadian Institute for Advanced Research). (2021). *AI & Healthcare: A Fusion of Law & Science — Regulation of Medical Devices with AI*. Toronto (ON): CIFAR.

CIHR, NSERC, & SSHRC (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada). (2016). *Tri-Agency Framework: Responsible Conduct of Research*. Ottawa (ON): CIHR, NSERC, SSHRC.

CIHR, NSERC, & SSHRC (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada). (2018a). *Tri-Agency Equity, Diversity and Inclusion Action Plan for 2018-2025*. Ottawa (ON): CIHR, NSERC, SSHRC.

CIHR, NSERC, & SSHRC (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council). (2018b). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Ottawa (ON): CIHR, NSERC, SSHRC.

CIHR (Canadian Institutes of Health Research). (2019). Tri-Agency Open Access Policy on Publications. Retrieved July 2021, from https://cihr-irsc.gc.ca/e/32005.html.

CIHR (Canadian Institutes of Health Research). (2021). Funding Overview. Retrieved February 2022, from https://cihr-irsc.gc.ca/e/37788.html.

CIPO (Canadian Intellectual Property Office). (2019). *IP Canada Report*. Ottawa (ON): Innovation, Science, and Economic Development Canada.

CIPO (Canadian Intellectual Property Office). (2021). What is a Patent? Retrieved June 2021, from https://www.ic.gc.ca/eic/site/cipointernet-internetopic.nsf/eng/wr03716.html#Section3.

Citron, D. K. & Pasquale, F. A. (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review, 89, 1-33*.

Ciuriak, D. & Ptashkina, M. (2021). *Quantifying Trade Secret Theft — Policy Implications*. Waterloo (ON): Centre for International Governance Innovation.

CleverHans. (2021). CleverHans. Retrieved August 2021, from https://github.com/cleverhans-lab/cleverhans.

cOAlition S. (2021). Principles and Implementation. Retrieved July 2021, from https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/.

CodeCarbon. (2021). CodeCarbon.io. Retrieved August 2021, from https://codecarbon.io/.

Cofone, I., Qarri, A., & Wiener, J. (2021, February 4). Updating Canada's Privacy Act for Artificial Intelligence, *Policy Magazine*.

Cowls, J., Tsamados, A., Floridi, L., & Taddeo, M. (2021). *The AI Gambit - Leveraging Artificial Intelligence to Combat Climate Change: Opportunities, Challenges, and Recommendations.* Berlin, Germany: Vodafone Institute for Society and Communications.

Craig, C. & Kerr, I. R. (2021). The death of the AI author. *Ottawa Law Review, 52*(1), 31–86.

Craig, C. J. (2021). AI and Copyright. In F. Martin-Bariteau & T. Scassa (Eds.), *Artificial Intelligence and the Law in Canada.* New York (NY): LexisNexis.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., . . . Whittaker, M. (2019). *AI Now 2019 Report.* New York (NY): AI Now Institute.

CSTI (Council for Science, Technology and Innovation). (2020). *Moonshot Goal 3.* Tokyo, Japan: Cabinet Office of Japan.

Cussins Newman, J. (2019). *Towards AI Security: Global Aspirations for a more Resilient Future.* Berkeley (CA): Center for Long-Term Cybersecurity.

Dastin, J. (2018, October 10). Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women, *Reuters.*

Davies, T., Walker, S. B., Rubinstein, M., & Perini, F. (Eds.). (2019). *The State of Open Data.* Cape Town, South Africa: African Minds and the International Development Research Centre (IDRC).

Dawson, P. (2021, May 23). ‘Ottawa Effect’ Could Unlock Benefits of Global Co-operation on AI, *The Globe and Mail.*

de Beer, J. (2020). *An International IP and Digital Trade Strategy for Canada.* Waterloo (ON): Centre for International Governance Innovation.

De Luna, P., Wei, J., Bengio, Y., Aspuru-Guzik, A., & Sargent, E. (2017). Use machine learning to find energy materials. *Nature, 552*(7683), 23–27.

Deep Genomics. (2021). AI-Powered Discovery. Retrieved September 2021, from https://www.deepgenomics.com/.

Deloitte Canada. (2019). *Canada’s AI Imperative — Overcoming Risks, Building Trust.* Toronto (ON): Deloitte LLP.

Deltorn, J.-M. (2017). Deep creations: Intellectual property and the automata. *Frontiers in Digital Humanities,* 4(3).

Dhar, P. (2020). The carbon impact of artificial intelligence. *Nature Machine Intelligence,* 2, 423–425.

Dick, S. (2019). Artificial intelligence. *Harvard Data Science Review*(1.1).

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology, 144*(1).

Dignum, V. (2020a). AI is multidisciplinary. *AI Matters, 5*(4), 18-21.

Dignum, V. (2020b). Responsibility and Artificial Intelligence. In M. D. Dubber, F. Pasquale & S. Das (Eds.), *The Oxford Handbook of Ethics of AI*. Oxford, United Kingdom: Oxford University Press.

Dilhac, M.-A., Abrassart, C., & Voarino, N. (Eds.). (2018). *Montréal Declaration for a Responsible Development of AI 2018*. Montréal (QC): Université de Montréal.

Dodds, L. & Wells, P. (2019). Data Infrastructure. In T. Davies, S. B. Walker, M. Rubinstein & F. Perini (Eds.), *The State of Open Data*. Cape Town, South Africa: African Minds and the International Development Research Centre (IDRC).

Données Québec. (2021). Organisations. Retrieved July 2021, from https://www.donneesquebec.ca/recherche/organization.

DORA (San Francisco Declaration on Research Assessment). (n.d.). San Francisco Declaration on Research Assessment. Retrieved September 2021, from https://sfdora.org/read/.

Douglas, H. (2015). Values in Science. In P. Humphreys (Ed.), *The Oxford Handbook of Philosophy of Science*. London, United Kingdom: Oxford University Press.

Dressel, J. & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*.

Egbuonu, K. (2021). The Latest News on the DABUS Patent Case. Retrieved October 2021, from https://www.ipstars.com/NewsAndAnalysis/the-latest-news-on-the-dabus-patent-case/Index/7366.

Element AI. (2019). *Global AI Talent Report 2019*. Montréal (QC): Element AI.

Element AI. (2020). *Global AI Talent Report 2020*. Montréal (QC): Element AI.

Engler, A. (2021). *How Open-Source Software Shapes AI Policy*. Washington (DC): Brookings Institution.

EOSC (European Open Science Cloud). (2021). *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)*. Brussels, Belgium: EOSC.

Eren Vural, I., Herder, M., & Graham, J. E. (2021). From sandbox to pandemic: Agile reform of Canadian drug regulation. *Health Policy, 125*(9), 1115-1120.

European Commission. (2018). *Evaluation of Directive 96/9/EC on the Legal Protection of Databases*. Brussels, Belgium: European Commission.

European Commission. (2021a). *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Brussels, Belgium: European Commission.

European Commission. (2021b). *Annexes to the Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Brussels, Belgium: European Commission.

European Parliament. (2019a). *Directive (EU) 2019/790 of the European Parliament and the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*. Brussels, Belgium: European Parliament.

European Parliament. (2019b). *Directive (EU) 2019/1024 of the European Parliament and the Council.* Brussels, Belgium: European Parliament.

European Parliament. (2020). European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence (2020/2014 (INL)). Retrieved July 2021, from https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html.

Expert Panel on Intellectual Property. (2020). *Report: Intellectual Property in Ontario's Innovation System*. Toronto (ON): Government of Ontario.

Extance, A. (2021, August 3). Machine Learning Delivers 'Human Genome' Moment for Proteins, *Chemistry World*.

Ezer, D. & Whitaker, K. (2019). Data science for the scientific life cycle. *Elife, 8*, e43979.

Fehr, C. (2011). What Is in It for Me? The Benefits of Diversity in Scientific Communities. In *Feminist Epistemology and Philosophy of Science*.

Fiddler, F. & Wilcox, J. (2018). Reproducibility of Scientific Results. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Summer 2021 ed.).

Fife, R. & Chase, S. (2021, July 12). Ottawa Imposes National Security Risk Assessments for University Researchers Seeking Federal Funds, *The Globe and Mail*.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI.* Cambridge (MA): Berkman Klein Center for Internet & Society.

Fleming, N. (2018). Computer-calculated compounds. *Nature, 557*, S55-S57.

FLI (Future of Life Institute). (2017a). Asilomar AI Principles. Retrieved September 2020, from https://futureoflife.org/ai-principles/.

FLI (Future of Life Institute). (2017b). A Principled AI Discussion in Asilomar. Retrieved September 2021, from https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/.

Floridi, L. (2021). The European Legislation on AI: A Brief Analysis of its Philosophical Approach. *Philosophy & Technology.*

Flynn, S., Geiger, C., Quintais, J., Margoni, T., Sag, M., Guibault, L., & Carroll, M. W. (2020). Implementing User Rights for Research in the Field of Artificial Intelligence. *European Intellectual Property Review*(7).

FNIGC (First Nations Information Governance Centre). (n.d.). About FNIGC. Retrieved August 2021, from https://fnigc.ca/about-fnigc/.

Fortin, J. M. & Currie, D. J. (2013). Big science vs. little science: How scientific impact scales with funding. *PLoS One, 8*(6), e65263.

FRDR (Federated Research Data Repository). (2021). About. Retrieved July 2021, from https://www.frdr-dfdr.ca/docs/en/about/.

Friesen, J. (2021, July 26). CSIS Warns Canadian Universities to be on Alert for International Espionage, *The Globe and Mail*.

Fruchterman, J. & Mellea, J. (2018). *Expanding Employment Success for People with Disabilities*. Benetech.

Gallini, N. & Hollis, A. (2019). *To Sell or Scale Up: Canada's Patent Strategy in a Knowledge Economy*. Montréal (QC): Institute for Research on Public Policy.

Gaudry, K. & Vandsburger, L. (2020). Determining the Likelihood that an AI patent application will be allowed at the USPTO. Retrieved July 2021, from https://www.ipwatchdog.com/2020/10/25/determining-likelihood-ai-patent-application-will-allowed-uspto/id=126687/.

GC (Government of Canada). (1985a). *Patent Act. R.S.C., 1985, c. P-4*. Ottawa (ON): GC.

GC (Government of Canada). (1985b). *Access to Information Act. R.S., 1985 c. A-1, s. 20*. Ottawa (ON): GC.

GC (Government of Canada). (2019). Intellectual Property Strategy. Retrieved April 2021, from https://www.ic.gc.ca/eic/site/108.nsf/eng/home.

GC (Government of Canada). (2020a). Open Data 101. Retrieved December 2021, from https://open.canada.ca/en/open-data-principles.

GC (Government of Canada). (2020b). Canada's Spatial Data Infrastructure. Retrieved August 2021, from https://www.nrcan.gc.ca/science-and-data/science-and-research/earth-sciences/geomatics/canadas-spatial-data-infrastructure/10783.

GC (Government of Canada). (2020c). *Annual Report of the Chief Science Advisor 2019-2020*. Ottawa (ON): Office of the Chief Science Advisor.

GC (Government of Canada). (2021a). The Government of Canada Launches Consultation on a Modern Copyright Framework for AI and the Internet of Things. Retrieved July 2021, from https://www.canada.ca/en/innovation-science-economic-development/news/2021/07/the-government-of-canada-launches-consultation-on-a-modern-copyright-framework-for-ai-and-the-internet-of-things.html.

GC (Government of Canada). (2021b). Open Data. Retrieved July 2021, from https://open.canada.ca/en/open-data.

GC (Government of Canada). (2021c). *Budget 2021 — A Recovery Plan for Jobs, Growth, and Resilience*. Ottawa (ON): Department of Finance Canada.

GC (Government of Canada). (2021d). Tri-Agency Research Data Management Policy. Retrieved July 2021, from https://science.gc.ca/eic/site/063.nsf/eng/h_97610.html.

GC (Government of Canada). (2021e). Model Policy on Scientific Integrity. Retrieved September 2021, from https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97643.html.

GC (Government of Canada). (2021f). Directive on Automated Decision-Making. Retrieved May 2021, from https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592&section=html.

GC (Government of Canada). (2021g). A Framework for Implementing Open-by-Default with Federal Government Science. Retrieved July 2021, from https://science.gc.ca/eic/site/063.nsf/eng/h_98201.html.

GDPR (General Data Protection Regulation). (2021). General Data Protection Regulation (GDPR). Retrieved August 2021, from https://gdpr.eu/article-89-processing-for-archiving-purposes-scientific-or-historical-research-purposes-or-statistical-purposes/.

Gebru, T., Morgenstern, J., Vecchione, B., Wortmann Vaughan, J., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *arXiv*(1803.09010v8).

Geist, M. (2021). AI and International Regulation. In F. Martin-Bariteau & T. Scassa (Eds.), *Artificial Intelligence and the Law in Canada*. Toronto (ON): LexisNexis.

Gherhes, C., Vorley, T., Vallance, P., & Brooks, C. (2021). The role of system-building agency in regional path creation: Insights from the emergence of artificial intelligence in Montreal. *Regional Studies*, 1-16.

Gibert, M., Mondin, C., & Chicoisne, G. (2018). Part 2: Overview of International Recommendations for AI Ethics. In M.-A. Dilhac, C. Abrassart & N. Voarino (Eds.), *Montréal Responsible AI Declaration*. Montréal (QC): Université de Montréal.

Gibert, M. (2020). *Faire la morale aux robots* (First ed.). Montreal (QC): Atelier 10.

GIDA (Global Indigenous Data Alliance). (n.d.-a). Purpose. Retrieved June 2021, from https://www.gida-global.org/purpose.

GIDA (Global Indigenous Data Alliance). (n.d.-b). Care Principles for Indigenous Data Governance. Retrieved June 2021, from https://www.gida-global.org/care.

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, 346(6206), 171-172.

Gil, Y., King, R. D., & Kitano, H. (2020). *AI Scientist Grand Challenge — Summary of Discussion During Workshop Held in February 2020*. London, United Kingdom: Alan Turing Institute.

Giuffrida, I. (2019). Liability for AI decision-making: Some legal and ethical considerations. *Fordham Law Review, 88*(2), 439–456.

Gov. of ON (Government of Ontario). (2017). Ontario Names First Chief Digital Officer. Retrieved September 2021, from https://news.ontario.ca/en/release/44109/ontario-names-first-chief-digital-officer.

Gov. of ON (Government of Ontario). (2021a). *Modernizing Privacy in Ontario*. Toronto (ON): Gov. of ON.

Gov. of ON (Government of Ontario). (2021b). Consultation: Ontario's Data Authority. Retrieved September 2021, from https://www.ontario.ca/page/consultation-ontarios-data-authority.

Gov. of ON (Government of Ontario). (2021c). Artificial Intelligence Guidance. Retrieved May 2021, from https://www.ontario.ca/page/artificial-intelligence-ai-guidance.

Gov. of the UK (Government of the United Kingdom). (2020). National Data Strategy. Retrieved July 2021, from https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy.

Grant, J. (2017, June 15). The Allocation of Scientific Grants Should be a Science, *Times Higher Education*.

GRC (Global Research Council). (2021). *Responsible Research Assessment — Global Research Council (GRC) Conference Report 2021*. GRC.

Gundersen, O. E., Gil, Y., & Aha, D. W. (2018). On reproducible AI: Towards reproducible research, open science, and digital scholarship in AI publications. *AI Magazine, 39*(3), 56-68.

Gundersen, O. E. & Kjensmo, S. (2018). *State of the Art: Reproducibility in Artificial Intelligence.* Paper presented at The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18).

Gunning, D. & Aha, D. W. (2017). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine, 40*(2), 44-58.

Guo, J., Farhang-Razi, V., & Algra, P. R. (2019). AI: A Glossary of Terms. In E. R. Erik R. Ranschaert, S. Sergey Morozov & P. R. Algra (Eds.), *Artificial Intelligence in Medical Imaging* (1 ed.). Cham, Switzerland: Springer.

Gupta, A., Lanteigne, C., & Kingsley, S. (2020). *A Social and Environmental Certificate for AI Systems*. Montréal (QC): Montreal AI Ethics Institute.

Habibollahi Najaf Abadi, H. & Pecht, M. (2020). Artificial Intelligence Trends Based on the Patents Granted by the United States Patent and Trademark Office. *IEEE Access, 8*, 81633-81643.

Haenlein, M. & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review, 61*(4), 5-14.

Hagen, G. D. (2021). AI and Patents and Trade Secrets. In F. Martin-Bariteau & T. Scassa (Eds.), *Artificial Intelligence and the Law in Canada*. Toronto (ON): LexisNexis.

HAI (Stanford University Human-Centered Artificial Intelligence). (2021). *Artificial Intelligence Index Report 2021*. Stanford (CA): Stanford University Human-Centered Artificial Intelligence.

Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Massive Analysis Quality Control (MAQC) Society Board of Directors, Waldron, L., . . . Aerts, H. J. W. L. (2020). Transparency and reproducibility in artificial intelligence. *Nature, 586*, E14–E16.

Hajian, S., Bonchi, F., & Castillo, C. (2016). *Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining.* Paper presented at 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco (CA).

Hatfield, P. W., Gaffney, J. A., Anderson, G. J., Ali, S., Antonelli, L., Basegmez du Pree, S., . . . Williams, B. (2021). The data-driven future of high-energy-density physics. *Nature, 593*(7859), 351-361.

HC (Health Canada). (2019). *Agile Regulations for Advanced Therapeutic Products and Clinical Trials*. Ottawa (ON): HC.

HC (Health Canada). (2021). Regulatory Innovation for Health Products: Enabling Advanced Therapeutic Products. Retrieved October 2021, from https://www.canada.ca/en/health-canada/corporate/about-health-canada/activities-responsibilities/strategies-initiatives/health-products-food-regulatory-modernization/advanced-therapeutic-products.html.

Heaven, W. D. (2018, November 22). AI peer reviewers unleashed to ease publishing grind, *Nature News.*

Heaven, W. D. (2020, November 12). AI is Wrestling with a Replication Crisis, *MIT Technology Review.*

Hecht, B., Wilcox, L., Bigham, J. P., Schöning, J., Hoque, E., Ernst, J., . . . Wu, C. (2018). It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. Retrieved June 2021, from https://acm-fca.org/2018/03/29/negativeimpacts/.

Hellström, T., Dignum, V., & Bensch, S. (2020). Bias in machine learning — what is it good for? *arXiv*(2004.00686v2).

Helmers, L., Horn, F., Biegler, F., Oppermann, T., & Muller, K. R. (2019). Automating the search for a patent's prior art with a full text similarity search. *PLoS One, 14*(3), e0212103.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). *Deep Reinforcement Learning that Matters.* Paper presented at The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans (LA).

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *arXiv.*

Hervey, M. (2020). Harnessing AI in drug discovery without losing patent protection. *Drug Discovery Today, 25*(6), 949-950.

Hinton, J. W. & Cowan, P. (2018, November 25). Canada's IP Strategy is Not in Step with Our Innovation and Commercialization Goals, *The Globe and Mail.*

Hinton, J. W. (2020, March 25). Canada Needs to Own Critical IP and Data Assets to Inspire Generational Economic Prosperity, *The Hill Times.*

Hocky, G. M. & White, A. D. (2022). Natural language processing models that automate programming will transform chemistry research and teaching. *arXiv*(2108.13360v2).

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The Diversity-innovation paradox in science. *Proceedings of the National Academy of Sciences of the United States of America, 117*(17), 9284-9291.

Hou, M., Ho, G., & Dunwoody, D. (2021). IMPACTS: A trust model for human-autonomy teaming. *Human-Intelligent Systems Integration, 3*(2), 79-97.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese, 169*, 615-626.

Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denuyl, S. (2020). Unintended Machine Learning Biases as Social Barriers for Persons with Disabilities. *ACM SIGACCESS Accessibility and Computing.*

Hutson, M. (2018, February 15). Missing Data Hinder Replication of Artificial Intelligence Studies, *Science Magazine.*

Hutson, M. (2021, April 1). AI conferences use AI to assign papers to reviewers. *Science Insider.*

ICPO (Information and Privacy Commissioner of Ontario). (2016). *De-identification Guidelines for Structured Data*. Toronto (ON): ICPO.

INDU (Standing Committee on Industry, Science and Technology). (2019). *Statutory Review of the Copyright Act*. Ottawa (ON): House of Commons Canada.

IPWatchdog. (2021). DABUS Gets its first Patent in South Africa under Formalities Examination. Retrieved July 2021, from https://www.ipwatchdog.com/2021/07/29/dabus-gets-first-patent-south-africa-formalities-examination/id=136116/.

Isdahl, R. & Gundersen, O. E. (2019). *Out-of-the-box Reproducibility: A Survey of Machine Learning Platforms.* Paper presented at 2019 IEEE 15th International Conference on e-Science and Grid Computing, San Diego (CA).

ISED (Industry, Science and Economic Development Canada). (2019a). *Processing Artificial Intelligence: Highlighting the Canadian Patent Landscape*. Ottawa (ON): Government of Canada.

ISED (Industry, Science and Economic Development Canada). (2019b). *Canada's Digital Charter in Action: A Plan by Canadians, for Canadians*. Ottawa (ON): Government of Canada.

ISED (Industry, Science and Economic Development Canada). (2021a). *A Consultation on a Modern Copyright Framework for AI and the Internet of Things*. Ottawa (ON): ISED.

ISED (Innovation, Science and Economic Development Canada). (2021b). Government of Canada takes action to protect Canadian research and intellectual property. Retrieved August 2021, from https://www.canada.ca/en/innovation-science-economic-development/news/2021/07/government-of-canada-takes-action-to-protect-canadian-research-and-intellectual-property.html.

ISED (Innovation, Science and Economic Development Canada). (2021c). Guidelines on the National Security Review of Investments. Retrieved April 2021, from https://www.ic.gc.ca/eic/site/ica-lic.nsf/eng/lk81190.html.

ITK (Inuit Tapiriit Kanatami). (n.d.). National Inuit Strategy on Research. Retrieved August 2021, from https://www.itk.ca/national-strategy-on-research-launched/.

Jia, X., Lynch, A., Huang, Y., Danielson, M., Lang'at, I., Milder, A., . . . Schrier, J. (2019). Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis. *Nature, 573*(7773), 251-255.

Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence, 1*(389), 389–399.

Johnson, L. (2021). Canada's AI Mafia. Retrieved May 2021, from https://xtract.ai/the-history-of-ai/.

Johnson, P. A., Sieber, R., Scassa, T., Stephens, M., & Robinson, P. (2017). The cost(s) of geospatial open data. *Transactions in GIS, 21*(3), 434-445.

JOURNE. (2021). JOURNE — Journal of Opportunities, Unexpected limitations, Retrospectives, Negative results, and Experiences. Retrieved August 2021, from https://sites.google.com/view/workshop-journe/home.

Kalvapalle, R. (2021, April 20). U of T Acceleration Consortium to use AI to develop advanced materials, *U of T News*.

Kelly, J. (2016). The text and data mining copyright exception - benefits and implications for UK higher education. Retrieved August 2021, from https://www.jisc.ac.uk/guides/text-and-data-mining-copyright-exception.

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., . . . Clare, A. (2009). The automation of science. *Science, 324*(5923), 85-89.

King, R. D. & Roberts, S. (2018). Chapter 5. Artificial Intelligence and Maching Learning in Science. In OECD (Ed.), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*. Paris, France: OECD Publishing.

King, R. D., Schüler-Costa, V., Mellingwood, C., & Soldatova, L. (2018). Automating sciences: Philosophical and social dimensions. *IEEE Technology and Society Magazine, 37*(1), 40–46.

Kitano, H. (2021). Nobel Turing Challenge: Creating the engine for scientific discovery. *NPJ Systems Biology and Applications*, 7(1), 29.

Knight, W. (2017, April 11). The Dark Secret at the Heart of AI, *MIT Technology Review.*

Knubley, J. (2021). *Building Superclusters for Canada.* Toronto (ON): Brookfield.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America, 111*(24), 8788-8790.

Kriegman, S., Blackiston, D., Levin, M., & Bongard, J. (2020). A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences of the United States of America, 117*(4), 1853-1859.

Kung, J., Boskovic, G., & Stix, C. (2020). *Building an AI World: Report on National and Regional AI Strategies.* Toronto (ON): Canadian Institute for Advanced Research.

Kusters, R., Misevic, D., Berry, H., Cully, A., Le Cunff, Y., Dandoy, L., . . . Wehbi, F. (2020). Interdisciplinary research in artificial Intelligence: Challenges and opportunities. *Frontiers in Big Data, 3*, 577974.

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning [pre-print]. *arXiv*(1910.09700v2).

Lamb, C. & Munro, D. (2020). *The Intangible Shift: Changing Gears to Compete in the New Economy.* Toronto (ON): The Brookfield Institute for Innovation + Entrepreneurship.

Lane, M. & Saint-Martin, A. (2021). *The Impact of Artificial Intelligence on the Labour Market: What do We Know so Far?* Paris, France: OECD Publishing.

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K.-R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications, 10*, 1096.

Larivière, V. & Sugimoto, C. (2018). Do authors comply with mandates for open access? *Nature, 562*, 483-486.

Laudel, G. (2006). The 'Quality Myth': Promoting and hindering conditions for acquiring research funds. *Higher Education, 52*, 375-403.

LCO (Law Commission of Ontario). (2021). *Regulating AI: Critical Issues and Choices*. Toronto (ON): Osgoode Law School, York University.

Leblond, P. (2019). *Digital Trade at the WTO — The CPTPP and CUSMA Pose Challenges to Canadian Data Regulation*. Waterloo (ON): Centre for International Governance Innovation.

Lehr, D. & Ohm, P. (2017). Playing with the data: What legal scholars should learn about machine learning. *UC Davis Law Review, 51*, 653-717.

Leonelli, S. (2018). Rethinking Reproducibility as a Criterion for Research Quality. In S. L. Fiorito, S. Scheall & C. Eduardo (Eds.), *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise* (Vol. 36B). Bingley, United Kingdom: Emerald Publishing Limited.

Leonelli, S. (2020). Scientific Research and Big Data. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 ed.).

Leonelli, S., Lovell, R., Wheeler, B. W., Fleming, L., & Williams, H. (2021). From FAIR data to fair data use: Methodological data fairness in health-related social media research. *Big Data & Society, 8*(1).

Lewis, J. E. (2020). *Indigenous Protocol and Artificial Intelligence Position Paper.* Honolulu (HI): The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research (CIFAR).

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., . . . Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data, 7*(1), 144.

Lipton, Z. C. (2018). The mythos of model interpretability. *Queue, 16*(3), 31–57.

Lo, B. & Field, M. J. (Eds.). (2009). *Conflict of Interest in Medical Research, Education, and Practice.* Washington (DC): The National Academies Press.

Local Contexts. (n.d.). About Us. Retrieved June 2021, from https://localcontexts.org/about/about-local-contexts/.

Logg, J. M., Minson, J. A., & Moore, D. A. (2018). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103.

Lopez-Gonzalez, J., Casalini, F., & Nemoto, T. (2021). Mapping Approaches to Cross-Border Data Flows. In I. Borchert & L. A. Winters (Eds.), *Addressing Impediments to Digital Trade*: CEPR Press.

Lopez, C. E., Miller, S. R., & Tucker, C. S. (2019). Exploring biases between human and machine generated designs. *Journal of Mechanical Design, 141*(2).

Lorsch, J. R. (2015). Maximizing the return on taxpayers' investments in fundamental biomedical research. *Molecular Biology of the Cell, 26*(9), 1578-1582.

Lowrie, I. (2017). Algorithmic rationality: Epistemology and efficiency in the data sciences. *Big Data & Society, 4*(1).

MacGregor, I. (2018, March 18). Big Data - The Canadian Opportunity, *CIGIOnline.*

Malone, M. (2020). Trade secrets, big data, and the future of public interest litigation over AI in Canada. *Canadian Intellectual Property Review, 35*, 6–9.

Marcia, V. & DeSouza, K. C. (2021, April 26). The EU Path Towards Regulation on AI, *Brookings TechTank.*

McKelvey, F. & Roberge, J. (2021, April 25). Canada is gambling with its leadership on artificial intelligence, *The Globe and Mail.*

Medeiros, M., Chalmers, W., & Sanft, J. (2021). Intellectual Property Strategy for AI Technology in Canada. In G. D'Agostino, A. Gaon & C. Piovesan (Eds.), *Leading Legal Disruption: Artificial Intelligence and a Tool Kit for Lawyers and the Law.* Toronto (ON): Thomson Reuters Canada.

Mercurio, L. & Yu, H. Y. (2021). An AI Policy for the (Near) Future. In I. Borchert & L. A. Winters (Eds.), *Addressing Impediments to Digital Trade*: CEPR Press.

Merton, R. (1968). The Matthew Effect in Science. *Science, 159*(3810), 56-63.

Metcalf, J. & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society, 3*(1).

Metcalf, J., Keller, E. F., & boyd, d. (2021). *Perspectives on Big Data, Ethics, and Society.* The Council for Big Data, Ethics, and Society.

Meyer, M. N. (2014). Misjudgements will drive social trials underground. *Nature, 511*, 265.

MGCS (Ministry of Government and Consumer Services). (2021). *Freedom of Information and Protection of Privacy Manual*. Ottawa (ON): Government of Ontario.

Mila (Quebec Artificial Intelligence Institute). (2021). Partnership Policy. Retrieved July 2021, from https://mila.quebec/en/mila/industrial-partnerships/partnership-policy/.

Mongeon, P., Brodeur, C., Beaudry, C., & Larivière, V. (2016). Concentration of research funding leads to decreasing marginal returns. *Research Evaluation, 25*(4), 396-404.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use, 37*(1), 49-56.

Montreal Declaration for Responsible AI. (2021). Signataires. Retrieved November 2021, from https://www.declarationmontreal-iaresponsable.com/signataires.

Moreley, J., Floridi, L., Kinsey, L. & Elhalal, A. (2021). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. In F. L. (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence. Philosophical Studies Series* (Vol. 144). Cham, Switzerland: Springer.

Morrison, O. (2019). Invention In the Age of Artificial Intelligence. What About Us? Retrieved September 2021, from https://medium.com/swlh/invention-in-the-age-of-artificial-intelligence-what-about-us-6fa4451c39a2.

Mourby, M., Gowans, H., Aidinlis, S., Smith, H., & Kaye, J. (2019). Governance of academic research data under the GDPR — Lessons from the UK. *International Data Privacy Law, 9*(3), 192–206.

Muggleton, S. H., Schmid, U., Zeller, C., Tamaddoni-Nezhad, A., & Besold, T. (2018). Ultra-Strong Machine Learning: Comprehensibility of programs learned with ILP. *Machine Learning, 107*(7), 1119-1140.

Muller, V. C. (2020). Ethics of ArtificiaI Intelligence and Robotics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 ed.).

Mulligan, C. & Elaluf-Calderwood, S. (2021). AI ethics: A framework for measuring embodied carbon in AI systems. *AI and Ethics*.

Mullin, R. (2021, March 28). The Lab of the Future is Now, *Chemical & Engineering News*.

Nagarajan, P., Warnell, G., & Stone, P. (2019). Deterministic implementations for reproducibility in deep reinforcement learning *arXiv*(1809.05676v5).

NAIIO (National Artificial Intelligence Initiative Office). (2021). Legislation and Executive Orders. Retrieved September 2021, from https://www.ai.gov/legislation-and-executive-orders/.

NASEM (National Academies of Sciences, Engineering, and Medicine). (2018). *Data Science: Opportunities to Transform Chemical Sciences and Engineering: Processing of a Workshop in Brief.* Washington (DC): NASEM.

NASEM (National Academies of Sciences, Engineering, and Medicine). (2019). *Reproducibility and Replicability in Science.* Washington (DC): The National Academies Press.

NASEM (National Academies of Sciences, Engineering, and Medicine). (2020). *Information Technology Innovation: Resurgence, Confluence, and Continuing Impact.* Washington (DC): The National Academies Press.

Naylor, C. D., Birgeneau, R. J., Crago, M., Lazaridis, M., Malacrida, C., McDonald, A. B., . . . Wilson, A. (2017). *Investing in Canada's Future: Strengthening the Foundations of Canadian Research*. Ottawa (ON): Canada's Fundamental Science Review.

NDRIO (New Digital Research Infrastructure Organization). (2021a). *Working Together to Create a Truly Integrated DRI Ecosystem in Canada.* Paper presented at 4th National Data Services Framework Summit, Virtual Summit.

NDRIO (New Digital Research Infrastructure Organization). (2021b). *Canadian DRI Needs Assessment – Preliminary Findings.* Toronto (ON): NDRIO.

NeurIPS (Conference on Neural Information Processing Systems). (2020). NeurIPS 2020 FAQ for Authors. Retrieved June 2021, from https://nips.cc/Conferences/2020/PaperInformation/NeurIPS-FAQ.

NeurIPS (Conference on Neural Information Processing Systems). (2021). NeurIPS 2021 Paper Checklist Guidelines. Retrieved June 2021, from https://neurips.cc/Conferences/2021/PaperInformation/PaperChecklist.

New Zealand Government. (2020). Government Chief Data Steward. Retrieved August 2021, from https://www.digital.govt.nz/digital-government/leadership/government-functional-leads/government-chief-data-steward-gcds/.

Nickles, T. (2018). Alien reasoning: Is a major change in scientific research underway? *Topoi, 39*(4), 901-914.

Niczepyr, D. (2021, January 12). National AI Initiative Office Launched by White House, *Fedscoop.*

Nielsen, M. W., Alegria, S., Börjeson, L., Etzkowitz, H., Falk-Krzesinski, H. J., Joshi, A., . . . Schiebinger, L. (2017). Opinion: Gender diversity leads to better science. *Proceedings of the National Academy of Sciences of the United States of America, 114*(13), E2796.

NIST (National Institute of Standards and Technology). (2020). High-Throughput Experimental Materials Collaboratory. Retrieved May 2021, from https://www.nist.gov/programs-projects/high-throughput-experimental-materials-collaboratory.

NRC (National Research Council). (2021). Artificial Intelligence for Design Challenge Program. Retrieved August 2021, from https://nrc.canada.ca/en/research-development/research-collaboration/programs/artificial-intelligence-design-challenge-program.

NRCan (Natural Resources Canada). (2019). *Canadian Geospatial Data Infrastructure (CGDI) — User Needs Assessment*. Ottawa (ON): NRCan.

NRCan (Natural Resources Canada). (2020). *Energy Fact Book 2020-2021*. Ottawa (ON): NRCan.

NSERC (National Sciences and Engineering Research Council). (2014). Draft Tri-Agency Open Access Policy – Frequently Asked Questions. Retrieved September 2021, from https://www.nserc-crsng.gc.ca/NSERC-CRSNG/policies-politiques/OpenAccessFAQ-LibreAccesFAQ_eng.asp.

O'Connor, A. M., Tsafnat, G., Thomas, J., Glasziou, P., Gilbert, S. B., & Hutton, B. (2019). A question of trust: Can we build an evidence base to gain trust in systematic review automation technologies? *Systematic Reviews, 8*(1), 143.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*, 477-453.

OECD (Organisation for Economic Co-operation and Development). (2018a). *Intelligence Machines, Smart Policies.* Paper presented at AI: Intelligent Machines, Smart Policies, Paris, France.

OECD (Organisation for Economic Co-operation and Development). (2018b). *OECD Science Technology and Innovation Outlook 2018 - Adapting to Technological and Social Disruption.* Paris, France: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development). (2019a). Recommendation of the Council on Artificial Intelligence. Retrieved September 2021, from https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

OECD (Organisation for Economic Co-operation and Development). (2019b). *Scoping the OECD AI Principles — Deliberations of the Expert Group on Artificial Intelligence at the OECD.* Paris, France: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development). (2021a). *Tools for Trustworthy AI.* Paris, France: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development). (2021b). *State of Implementation of the OECD AI Principles — Insights from National AI Policies*. Paris, France: OECD Publishing.

OECD (Organisation for Economic Co-Operation and Development). (2021c). *Recommendation of the Council on Artificial Intelligence.* Paris, France, OECD Publishing.

OECD (Organisation for Economic Co-operation and Development). (2021d). Shaping an enabling policy environment for AI (Principle 2.3). Retrieved August 2021, from https://www.oecd.ai/dashboards/ai-principles/P12.

OECD (OECD.AI Policy Observatory). (2021e). *OECD Metrics & Methods.* Paris, France: OECD Publishing.

OHDP (Ontario Health Data Platform). (2020). About OHDP. Retrieved August 2021, from https://ohdp.ca/overview/.

OPC (Office of the Privacy Commissioner of Canada). (2021a). Summary of Privacy Laws in Canada. Retrieved May 2021, from https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/.

OPC (Office of the Privacy Commissioner of Canada). (2021b). *Police Use of Facial Recognition Technology in Canada and the Way Forward: Special Report to Parliament on the OPC's Investigation into the RCMP's Use of Clearview AI and Draft Joint Guidance for Law Enforcement Agencies Considering the Use of Facial Recognition Technology.* Ottawa (ON): OPC.

OPC (Office of the Privacy Commissioner of Canada). (2021c). *Joint investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta*. Ottawa (ON): OPC.

Oransky, I. & Marcus, A. (2017, July 20). Peer Review Is A Black Box. Let's Open It Up, *Stat*.

Parliament of Canada. (2020). *An Act to Implement the Agreement Between Canada, the United States of America and the United Mexican States*. Ottawa (ON): Parliament of Canada.

Partnership on AI. (2021). *Managing the Risks of AI Research – Six Recommendations for Responsible Publication.* San Francisco (CA): Partnership on AI.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., . . . Dean, J. (2021). Carbon Emissions and Large Neural Network Training [pre-print]. *arXiv*(2104.10350v3).

Pritt, M. (2018). Between a Rock and a Hard Place: Protecting Employee Mobility and Negative Trade Secrets. Retrieved July 2021, from https://www.bsfllp.com/news-events/between-a-rock-and-a-hard-place-protecting-employee-mobility-and-negative-trade-secrets.html.

Rainie, S. C., Rodriguez-Lonebear, D., & Martinez, A. (2017). *Indigenous Data Sovereignty in the United States.* Tuscon (AZ): University of Arizona Natives Nations Institute.

Reardon, S. (2015). Canadians baulk at reforms to health-research agency. *Nature, 520*(7547), 272-273.

Renda, A. (2019). *Artificial Intelligence — Ethics, Governance, and Policy Challenges*. Brussels, Belgium: Centre for European Policy Studies (CEPS).

Reynolds, A. & Lewis, D. (2017, March 30). Teams Solve Problems Faster When They're More Cognitively Diverse, *Harvard Business Review.*

Reznick, R. K., Harris, K., & Horsley, T. (2020). *Task Force Report on Artificial Intelligence and Emerging Digital Technologies*. Ottawa (ON): Royal College of Physicians and Surgeons of Canada.

Roberts, H., Cowls, J., Hine, E., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). China's artificial intelligence strategy: Lessons from the European Union's 'ethics-first' approach. *Social Science Research Network*.

Roberts, J. & Fisher, D. (2020). *pReview: The Artificially Intelligent Conference Reviewer.* Paper presented at 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami (FL).

Rolin, K. (2019). The Epistemic Significance of Diversity. In M. Fricker, P. J. Graham, D. Henderson & N. J. L. L. Pedersen (Eds.), *The Routledge Handbook of Social Epistemology.* New York (NY): Routledge.

RRITools. (n.d.). *A Practical Guide to Responsible Research and Innovation*. Valencia, Spain: La Caixa Foundation.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206-215.

Rudin, C. & Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*(1.2).

Samek, W. & Müller, K.-R. (2019). Towards Explainable Artificial Intelligence. *arXiv*(1909.12072v1).

Sanderson, K. (2019). Automation: Chemistry shoots for the moon. *Nature, 568*(577–579).

Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). *The Risk of Racial Bias in Hate Speech Detection.* Paper presented at 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy.

Satariano, A. (2020, August 20). British Grading Debacle Shows Pitfalls of Automating Government, *The New York Times*.

Scassa, T. (2019a). Ownership and control over publicly accessible platform data. *Online Information Review, 43*(6), 986-1002.

Scassa, T. (2019b). Privacy. In T. Davies, S. B. Walker, M. Rubinstein & F. Perini (Eds.), *The State of Open Data*. Cape Town, South Africa: African Minds and the International Development Research Centre (IDRC).

Scassa, T. (2021a). Legal Issues Around Data Scraping as a Source of Data for AI Innovation. In G. D'Agostino, A. Gaon & C. Piovesan (Eds.), *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law.* Toronto (ON): Thomson Reuters Canada.

Scassa, T. (2021b). Rights in Data, the Public Interest, and International Trade Law. In I. Borchert & L. A. Winters (Eds.), *Addressing Impediments to Digital Trade.* London, United Kingdom: Centre for Economic Policy Research Press.

Scassa, T. (2021c). AI and Data Protection Law. In F. Martin-Bariteau & T. Scassa (Eds.), *Artificial Intelligence and the Law in Canada*. Montréal (QC): LexisNexis.

SCC (Standards Council of Canada). (2020). Canadian Data Governance Standardization Collaborative. Retrieved May 2021, from https://www.scc.ca/en/flagships/data-governance.

Schmidt, V., Luccioni, A., Lacoste, A., & Dandres, T. (n.d.). Machine Learning CO_2 Impact Calculator. Retrieved December 2021, from https://mlco2.github.io/impact/.

Schwartz, O. (2019, November 25). In 2016, Microsoft's Racist Chatbot Revealed the Danger of Online Conversation, *IEEE Spectrum*.

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the Association for Computing Machinery, 62*(12), 54-63.

Science International. (2015). *Open Data in a Big Data World*. Paris, France: International Council for Science (ICSU), International Social Science Council (ISSC), The World Academy of Sciences (TWAS), InterAcademy Partnership.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal, 314*, 498-513.

Selbst, A. D. & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. *Fordham Law Review, 87*, 1085-1139.

Sethi, A., Sankaran, A., Panwar, N., Khare, S., & Mani, S. (2018). *DLPaper2Code Auto-Generation of Code from Deep Learning Research Papers.* Paper presented at The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans (LA).

Simonite, T. (2018, August 17). AI is the Future—But Where Are the Women?, *WIRED.*

Smith, D. (2021, August 17). Here's What Died on the Order Paper, *National Magazine*.

Smith, M. (2017, December 6). How Canada has emerged as a leader in artificial intelligence, *University Affairs.*

Somers, J. (2021, March 18). The Pastry AI that Learned to Fight Cancer, *The New Yorker.*

Sparkes, A., Aubrey, W., Byrne, E., Clare, A., Khan, M. N., Liakata, M., . . . King, R. D. (2010). Towards robot scientists for autonomous scientific discovery. *Automated Experimentation, 2*, 1.

Sparkes, M. (2021, July 15). World's first 3D printed steel bridge opens in Amsterdam, *New Scientist.*

SSHRC (Social Sciences and Humanities Research Council). (2021). New Frontiers in Research Fund. Retrieved October 2021, from https://www.sshrc-crsh.gc.ca/funding-financement/nfrf-fnfr/index-eng.aspx.

Stach, E., DeCost, B., Kusne, A. G., Hattrick-Simpers, J., Brown, K. A., Reyes, K. G., . . . Maruyama, B. (2021). Autonomous experimentation systems for materials development: A community perspective. *Matter, 4*(9), 2702-2726.

StatCan (Statistics Canada). (2016). *Occupation – National Occupational Classification (NOC) 2016 (693A), Highest Certificate, Diploma or Degree (15), Labour Force Status (3), Age (13A) and Sex (3) for the Labour Force Aged 15 Years and Over in Private Households of Canada, Provinces and Territories, Census Metropolitan Areas and Census Agglomerations, 2016 Census – 25% Sample Data*. Ottawa (ON): StatCan.

StatCan (Statistics Canada). (2021a). *Access to the Internet in Canada 2020 – June 2021.* Ottawa (ON): StatCan.

StatCan (Statistics Canada). (2021b). *Access to the Internet in Canada 2020 – May 2021.* Ottawa (ON): StatCan.

Steffler, J. (2016). The Indigenous Data Landscape in Canada: An Overview. *aboriginal policy studies, 5*(2).

Stevens, R., Nichols, J., & Yelick, K. (2020). *AI for Science*. Washington (DC): U.S. Department of Energy.

Stoyanovich, J., Abiteboul, S., & Miklau., G. (2016). *Data, Responsibly: Fairness, Neutrality and Transparency in Data Analysis.* Paper presented at International Conference on Extending Database Technology, Bordeaux, France.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv*(1906.02243v1).

Taddeo, M., Tsamados, A., Cowls, J., & Floridi, L. (2021). Artificial intelligence and the climate emergency: Opportunities, challenges, and recommendations. *One Earth, 4*(6), 776-779.

Taylor, J. (2021, July 30). I'm sorry Dave I'm afraid I invented that: Australian court finds AI systems can be recognised under patent law, *The Guardian.*

The Turing Institute (The Alan Turing Institute). (2021). *A Roadmap for Developing AI Systems Capable of Making Nobel Quality Scientific Discoveries*. London, United Kingdom: The Alan Turing Institute.

The Alliance (Digital Research Alliance of Canada). (2021). Introducing the Digital Research Alliance of Canada. Retrieved October 2021, from https://engagedri.ca/latest/introducing-the-digital-research-alliance-of-canada.

The Artificial Inventor Project. (2021). Home. Retrieved September 2021, from https://artificialinventor.com/.

The Royal Society. (2017). *Machine Learning: The Power and Promise of Computers that Learn by Example.* London, United Kingdom: The Royal Society.

The Royal Society. (2019). *The AI Revolution in Scientific Research.* London, United Kingdom: The Royal Society.

Thompson, E. (2021, February 3). U.S. Technology Company Clearview AI Violated Canadian Privacy Law: Report, *CBC News.*

Thompson, K. & Lucarini, L. (2021). CPPA: An in-depth look at the de-identification provisions in Canada's proposed new privacy law. Retrieved February 2022, from http://www.dentonsdata.com/cppa-an-in-depth-look-at-the-de-identification-provisions-in-canadas-proposed-new-privacy-law/.

TMS (The Minerals, Metals & Materials Society). (2017). *Building a Materials Data Infrastructure — Opening New Pathways to Discovery and Innovation in Science and Engineering*. Pittsburgh (PA): TMS.

Tortoise Media. (2020). *The Global AI Index — Methodology*. London, United Kingdom: Tortoise Media.

Tortoise Media. (2021). Global AI Index Rankings. Retrieved May 2021, from https://www.tortoisemedia.com/intelligence/global-ai/.

UNESCO (United Nations Educational, Scientific and Cultural Organization). (2018). Canada First to Adopt Strategy for Artificial Intelligence. Retrieved May 2021, from http://www.unesco.org/new/en/media-services/single-view/news/canada_first_to_adopt_strategy_for_artificial_intelligence/.

UNESCO (United Nations Educational, Scientific and Cultural Organization). (2020). *Outcome Document: First Draft of the Recommendation on the Ethics of Artificial Intelligence*. Paris, France: UNESCO.

UofT (University of Toronto). (2021a). Indigenous Data Sovereignty. Retrieved December 2021, from https://guides.library.utoronto.ca/indigenousstudies/datasovereignty.

UofT (University of Toronto). (2021b). Pilot Program Embeds Ethics Into U of T Undergraduate Technology Courses. Retrieved August 2021, from https://www.utoronto.ca/news/pilot-program-embeds-ethics-u-t-undergraduate-technology-courses.

U.S. NRC (U.S. National Research Council). (2014). *Big Data in Materials Research and Development: Summary of a Workshop*. Washington (DC): The National Academies Press.

USIDSN (United States Indigenous Data Sovereignty Network). (n.d.). Join The US Indigenous Data Sovereignty Network. Retrieved October 2021, from https://usindigenousdata.org/join.

van der Aalst, W. M. P., Bichler, M., & Heinzl, A. (2017). Responsible data science. *Business & Information Systems Engineering, 59*(5), 311–313.

Van Roy, V. (2019). *AI Watch National Strategies on Artificial Intelligence — A European Perspective in 2019*. Luxembourg, Luxembourg: European Commission.

Vayena, E. & Blasimme, A. (2017). Biomedical Big Data: New Models of Control Over Access, Use and Governance. *Journal of Bioethical Inquiry, 14*(4), 501–513.

Veale, M. & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society, 4*(2).

Vector Institute. (n.d.). *Intellectual Property Policy*. Toronto (ON): Vector Institute.

Verma, I. M. (2014). Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America, 111*(29), 10779.

Villeneuve, S., Boskovic, G., & Barron, B. (2019). *Rebooting Regulation: Exploring the Future of AI Policy in Canada*. Toronto (ON): Canadian Institute for Advanced Research.

Wainberg, M., Merico, D., Delong, A., & Frey, B. J. (2018). Deep learning in biomedicine. *Nature Biotechnology, 36*(9), 829-838.

Wang, D. & Barabási, A.-L. (2021). *The Science of Science* (1st ed.). Cambridge, United Kingdom: Cambridge University Press.

Wang, W. Y., Tang, B., Lin, D. Y., Zou, C. X., Zhang, Y., Shang, S. L., . . . Li, J. S. (2020). A brief review of data-driven ICME for intelligently discovering advanced structural metal materials: Insight into atomic and electronic building blocks. *Journal of Materials Research, 35*(8), 872-889.

Web Foundation (World Wide Web Foundation). (2018). *Open Data Barometer – Leaders Edition*. Washington (DC): World Wide Web Foundation.

Webb, M. (2020). The Impact of Artificial Intelligence on the Labor Market. *Social Science Research Network*.

Webster, P. (2015). CIHR modifies virtual peer review amidst complaints. *Canadian Medical Association Journal, 187*(5), E151-E152.

WEF (World Economic Forum). (2018). *How to Prevent Discriminatory Outcomes in Machine Learning*. WEF.

Weis, J. W. & Jacobson, J. M. (2021). Learning on knowledge graph dynamics provides an early warning of impactful research. *Nature Biotechnology, 39*, 1300–1307.

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating Systems: Gender, Race, and Power in AI*. AI NOW Institute.

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and Societal Implications Algorithms, Data, and Artificial Intelligence: A Roadmap for Research*. London, United Kingdom: Nuffield Foundation.

Whittlestone, J. & Ovadya, A. (2020). The tension between openness and prudence in responsible AI research. *arXiv*(1910.01170v2).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018.

WIPO (World Intellectual Property Organization). (2020). *WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI)*. Geneva, Switzerland: WIPO.

Wolchover, N. (2018, April 18). Machine Learning's 'Amazing' Ability to Predict Chaos, *Quanta Magazine*.

Woodgett, J. (2018). Peer Review and the Real Cost of Popularism (or How to Get 640 Comments on an Article). Retrieved August 2021, from https://jwoodgett.medium.com/peer-review-and-the-real-cost-of-popularism-or-how-to-get-640-comments-on-an-article-d9522d4cd668.

Woolston, C. & Perkel, J. M. (2020). Sliced, diced and digested: Science ready in minutes. *Nature, 588*, S136–S137.

WTO (World Trade Organization). (2017). *Agreement on Trade-Related Aspects of Intellectual Property Rights (as Amended on 23 January 2017).* Geneva, Switzerland: WTO.

Wu, B., Maxwell, T., & Crew, B. (2020). Network effect: visualizing AI connections in the natural sciences. *Nature*(588), S112–S113.

Wylie, B. (2018). *Open Data Endgame - Countering the Digital Consensus*. Waterloo (ON): Centre for International Governance Innovation.

Xinhuanet. (2020). China Promotes Local AI Pilot Zones. Retrieved September 2021, from http://www.xinhuanet.com/english/2020-05/09/c_139043630.htm.

Xue, M., Yuan, C., Wu, H., Zhang, Y., & Liu, W. (2020). Machine Learning Security: Threats, Countermeasures, and Evaluations. *IEEE Access, 8*, 74720-74742.

Yeston, J. (2021). Progress in Data and Code Deposition. Retrieved July 2021, from https://blogs.sciencemag.org/editors-blog/2021/07/15/progress-in-data-and-code-deposition/.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series, 1168*, 022022.

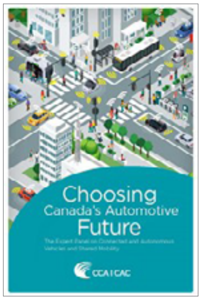
Young, E., Wajcman, J., & Sprejer, L. (2021). *Where Are the Women? Mapping the Gender Job Gap in AI*. The Alan Turing Institute.

Yu, P. K. (2020). The algorithmic divide and equality in the age of artificial intelligence. *Florida Law Review, 72*.

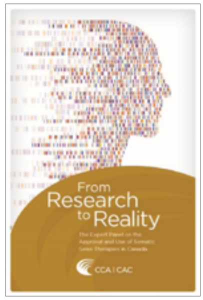
CCA Reports of Interest

The assessment reports listed below are available on the CCA's website (www.cca-reports.ca):

Powering Discovery (2021)



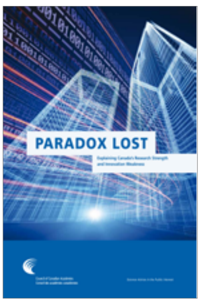
Choosing Canada's Automotive Future (2021)



From Research to Reality (2020)

Building Excellence (2019)

Competing in a Global Innovation Economy: The Current State of R&D in Canada (2018)



Paradox Lost: Explaining Canada's Research Strength and Innovation Weakness (2013)

Strengthening Canada's Research Capacity: The Gender Dimension (2012)

Honesty, Accountability and Trust: Fostering Research Integrity in Canada (2010)

CCA Board of Directors*

*As of January 2022

CCA Scientific Advisory Committee*

David Castle (Chair), Professor, School of Public Administration and Gustavson School of Business, University of Victoria; Researcher in Residence, Office of the Chief Science Advisor to the Prime Minister of Canada (Victoria, BC)

Maydianne C. B. Andrade, Professor, Biological Sciences, University of Toronto Scarborough; President, Canadian Black Scientists Network (Toronto, ON)

Peter Backx, FRSC, Professor, Department of Biology, and Canada Research Chair in Cardiovascular Biology, York University (Toronto, ON)

Stephanie E. Chang, Professor, School of Community and Regional Planning and Institute for Resources, Environment and Sustainability, University of British Columbia (Vancouver, BC)

Neena L. Chappell, C.M., FRSC, FCAHS, Professor Emeritus, Institute on Aging and Lifelong Health and the Department of Sociology, University of Victoria (Victoria, BC)

Jackie Dawson, Canada Research Chair in Environment, Society and Policy, and Associate Professor, Department of Geography, University of Ottawa (Ottawa, ON)

Colleen M. Flood, FRSC, FCAHS, Director, Centre for Health Law, Policy and Ethics; Professor, Faculty of Law (Common Law Section), University of Ottawa (Ottawa, ON)

Digvir S. Jayas, O.C., FRSC, FCAE, Distinguished Professor and Vice-President (Research and International), University of Manitoba (Winnipeg, MB)

Malcolm King, FCAHS, Scientific Director, Saskatchewan Centre for Patient-Oriented Research, University of Saskatchewan (Saskatoon, SK)

Chris MacDonald, Associate Professor; Director, Ted Rogers Leadership Centre; Chair, Law and Business Department; Ted Rogers School of Management, Ryerson University (Toronto, ON)

Barbara Neis, C.M., FRSC, John Paton Lewis Distinguished University Professor, Memorial University of Newfoundland (St. John's, NL)

Nicole A. Poirier, FCAE, President, KoanTeknico Solutions Inc. (Beaconsfield, QC)

Jamie Snook, Executive Director, Torngat Wildlife Plants and Fisheries Secretariat (Happy Valley-Goose Bay, NL)

David A. Wolfe, Professor of Political Science, University of Toronto Mississauga; Co-Director, Innovation Policy Lab, Munk School of Global Affairs and Public Policy, University of Toronto (Toronto, ON)

We were deeply saddened to learn of the passing of **Jeffrey A. Hutchings, FRSC,** Killam Memorial Chair and Professor of Biology, Dalhousie University, who was a valued member of the Scientific Advisory Committee. Our deepest sympathies go out to his family, friends, and colleagues.

*As of January 2022

Council of Canadian Academies | Conseil des académies canadiennes

180 Elgin Street, Suite 1401
Ottawa ON K2P 2K3
Tel: 613 567-5000
www.cca-reports.ca