

# Reimagining Regulation for the Age of AI: New Zealand Pilot Project

WHITE PAPER

JUNE 2020



# Contents

3	Introduction
4	1 Project overview
8	2 Reimagining AI regulation
13	3 Focus areas
14	3.1 National conversation
17	3.2 Regulatory capabilities and institutional design
20	3.3 Risk/benefit assessment of AI systems for government
24	Conclusion
25	Contributors
27	Endnotes

© 2020 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

# Introduction

Over the past decade, artificial intelligence (AI) has emerged as the software engine that drives the Fourth Industrial Revolution, a technological force affecting all disciplines, economies and industries. AI-powered services are already being applied to create more personalized shopping experiences,<sup>1</sup> drive productivity<sup>2</sup> and increase farming efficiency. In the future, they will enable the rise of self-driving cars<sup>3</sup> and the large-scale access to precision medicine with appropriate data governance.<sup>4</sup> AI systems have been able to do so thanks to the exponential growth of human and machine-generated data leveraged by powerful machine learning algorithms,<sup>5</sup> whose performance on a given task increases with labelled data.

This recent progress is remarkable in important respects, but also creates unique challenges. Indeed, without proper oversight, AI may replicate or even exacerbate<sup>6</sup> human bias and discrimination, cause potential job displacement<sup>7</sup> and lead to other unintended<sup>8</sup> consequences. This is particularly problematic when AI is deployed in high-stakes domains such as criminal justice,<sup>9</sup> healthcare,<sup>10</sup> banking<sup>11</sup> or employment.<sup>12</sup>

Government officials throughout the world are increasingly aware of both the opportunities and risks associated with AI and urged to act as AI's influence over society increases at a fast pace. They also acknowledge that some form of AI regulation is needed, with AI systems used by governments an early focus, given the duty of care owed to citizens, particularly as governments make highly consequential decisions supported by AI.

Yet, regulating AI is a complex endeavour. Experts hold diverse views on what areas and activities should be regulated, and approaches to regulating

AI diverge sharply across regions. In some jurisdictions, a lack of consensus on a path forward and the risk of stifling innovation may deter any action. Emerging controversies surrounding AI can also force governments to implement hastily constructed and suboptimal regulatory policies. What is possible, however, is to start to address some of the key issues in AI through tangible solutions and tools that could be leveraged by national governments.

To this end, the World Economic Forum is spearheading a multistakeholder, evidence-based policy project in partnership with the Government of New Zealand. The project aims at co-designing actionable governance frameworks for AI regulation. It is structured around three focus areas: 1) obtaining of a social licence for the use of AI through an inclusive national conversation; 2) the development of in-house understanding of AI to produce well-informed policies; and 3) the effective mitigation of risks associated with AI systems to maximize their benefits.

For each of these areas, the World Economic Forum's Centre for the Fourth Industrial Revolution – under the Platform for Shaping the Future of Technology Governance: Artificial Intelligence and Machine Learning – has produced frameworks and guidelines that, once combined, contribute to the development of an appropriate regulatory environment for AI.

In the next stage, the Centre intends to pilot the frameworks to develop a better understanding of what works and why. This white paper is the first step in an iterative process, and welcomes participation of organizations willing to engage in this debate and join the project.

1

# Project overview

This is the first global multistakeholder effort to co-design regulatory frameworks for AI informed by policy pilots



“ Having different voices and viewpoints in conversations and decisions is vital to stop the risk of creating new digital divides.

The World Economic Forum Platform for Shaping the Future of Technology Governance: Artificial Intelligence and Machine Learning is guided by a vision of accelerating and scaling the social benefits of emerging technologies while mitigating their risks. To execute this vision, the Platform – which is guided by the Forum’s Centre for the Fourth Industrial Revolution – applies an agile governance methodology to develop actionable frameworks or toolkits through a multistakeholder approach that brings together governments, companies, civil society and academia. The Centre partners with governments and companies to pilot the frameworks, capture lessons and create guidance to support broad scaling and self-service adoption.

The Platform develops projects within three distinctive areas: enabling frameworks, high-risk use cases and leapfrog opportunities. Projects in the first category – which includes Reimagining Regulation for the Age of AI – aim to create enabling frameworks that support the operationalization of the ethical use of artificial intelligence. Earlier projects have done this in a variety of contexts, including government procurement, corporate operational management and corporate governance. This project complements the existing portfolio by addressing the potential need for the regulation of AI in some form.

## Working with the New Zealand government

New Zealand is the sponsor government for this project. A number of initiatives in New Zealand – such as the government’s Algorithm Assessment Report,<sup>13</sup> the Centre for AI and Public Policy, Otago University report, Government Use of AI in New Zealand,<sup>14</sup> and the AI Forum of New Zealand’s work on AI in the economy and society<sup>15</sup> – have raised the importance of AI and explored opportunities, but the government has not yet developed an AI strategy.

This project is seen as an opportunity to work with leading experts in AI to help New Zealand shape its domestic position on emerging technologies. New Zealand has expressed interest in working with the Centre for the Fourth Industrial Revolution on this topic, given the need for a global, multistakeholder perspective on the complex question of regulating AI. New Zealand has been keen to work with the Centre to identify tools and approaches that would promote innovation, protect society and build trust in AI use. Its view is that ethical frameworks need to underpin technology development, and New Zealand’s bicultural foundation and multicultural make-up means it is committed to working with communities to build and maintain social licence for the use of technologies.

In this paper, social licence is defined as people giving approval to an organization they deem

trustworthy enough to undertake specified, potentially risky activities. This trust needs to be gained by building constructive relationships with stakeholders.<sup>16</sup> Key to gaining social licence is demonstrating that respect for human rights and the willingness to operate ethically informs the work being done. Social licence is clearly highly contextual to a specific nation, state, or community.

As an example, within New Zealand, nationwide discussions will have to involve New Zealand’s indigenous people, the Māori, and their worldview that everything living and non-living is interconnected. People act as kaitiaki (guardians) to preserve the land and everything on it, including intangible items such as data. Building trust and gaining social licence within a New Zealand context must recognize this Te Ao Māori worldview and have Te Tiriti o Waitangi (New Zealand’s founding document) and its principles at its heart.<sup>17</sup> Having different voices and viewpoints in conversations and decisions is vital to stop the risk of creating new digital divides. It will also be important to prioritize people’s well-being, in line with the government’s commitment to improving the well-being of all New Zealanders. This work is focused on promoting higher living standards and greater intergenerational well-being for New Zealanders, making sure the four capitals of human, social, natural and financial/physical are strong and working well together.<sup>18,19</sup>

## Project timeline

The project schedule is as follows:

- Scoping (*September to December 2019*): Build core project community of key stakeholders and identify primary issues and knowledge base
- Co-designing (*January to June 2020*): Work with the project community to frame the conversation on AI regulation, identify focus areas, draft governance frameworks and select pilot projects
- Testing (*July to December 2020*): Pilot new approaches and tools for AI regulation and capture lessons and share findings
- Scaling (*January 2021 onward*): Encourage broad adoption of the designed governance frameworks and tools based on lessons learned from pilot implementations

## Key project milestones

Two multistakeholder workshops were organized by the project team: the first in Wellington, New Zealand in October 2019 with a New Zealand-based community; and a second with a global community in San Francisco in January 2020. The conversations were not at a technical level, reflecting the interests of the multistakeholder group, and in line with the Centre for the Fourth Industrial Revolution's focus on agile governance methodology. More detail in particular areas was discussed in a series of webinars and conversations between the project team and the global community following the workshops.

What was clear from the workshops was that trust cannot be built without having strong engagement

and transparent models and systems that are seen to be safe, reliable and respectful of human rights. Gaining this trust requires using collaborative and inclusive tools and approaches that allow people from all walks of life to give their views.

Further planned workshops and engagement with the New Zealand and global communities were cancelled due to COVID-19. Instead, a series of webinars were held to work with the communities on the development of this white paper. It is hoped that a certain amount of pilot project activity will continue in New Zealand, and that the process of engaging other pilot partners can restart in earnest when the course of the global pandemic has become more predictable and its effects have diminished.

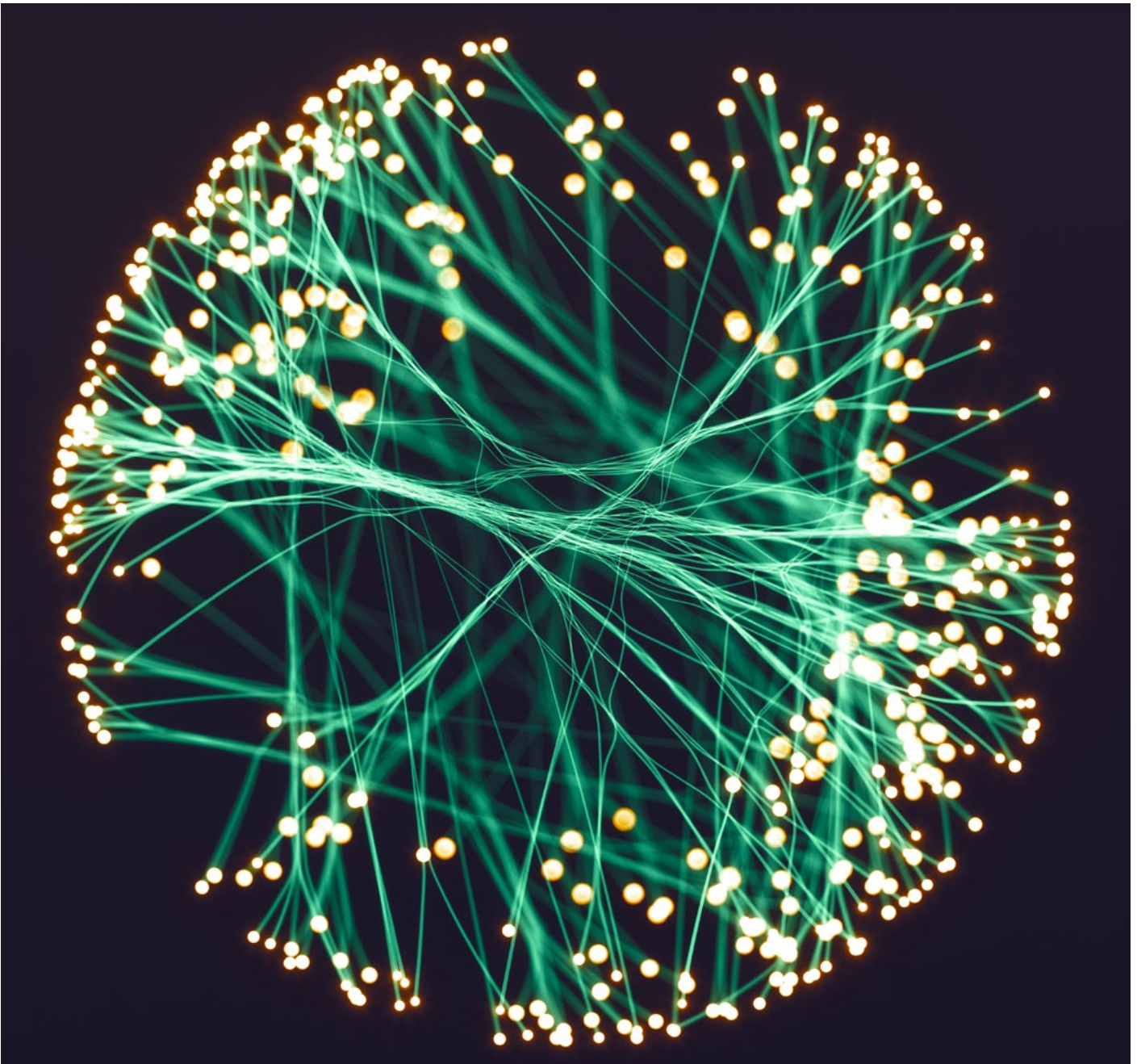


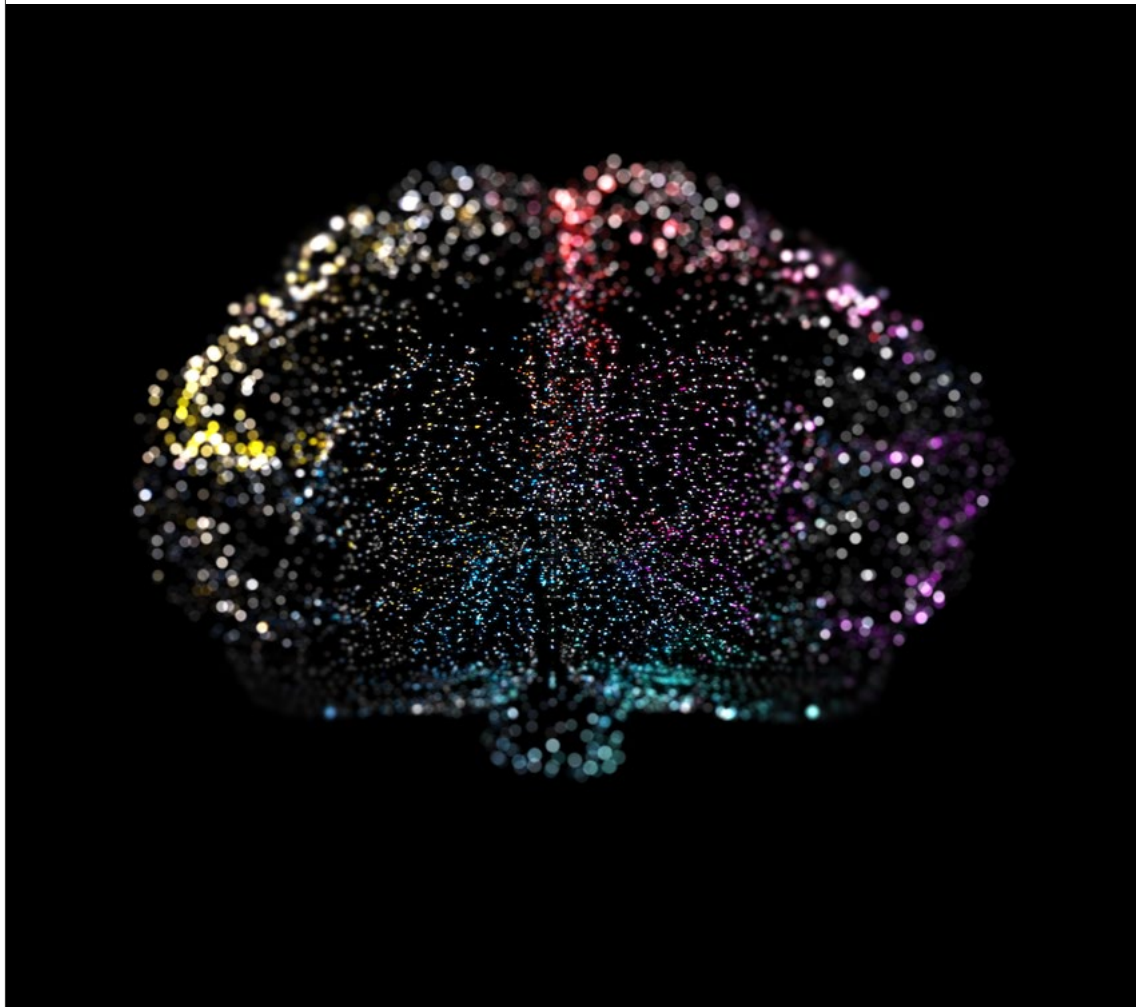
FIGURE 1 | Workshop insights

New Zealand workshop (October 2019)	Global workshop in San Francisco (January 2020)
<p><b>Focus</b></p> <hr/> <p>“How do we keep our citizens safe in the age of artificial intelligence?”</p> <p><b>Key takeaways</b></p> <hr/> <ul style="list-style-type: none"> <li>– Recognize that transparency, in all its meanings, is a core value (for example, people want to know if the models have been thoroughly tested, why particular decisions have been made, what the outcomes will be seen from the use of AI)</li> <li>– Build trust, which is necessary for AI to be fully accepted and requires being inclusive by inviting and respecting the views of all parts of the community</li> <li>– Use a people-centred design (e.g. co-design with Māori in New Zealand)</li> <li>– Ensure human rights underpins the work</li> <li>– Include business in conversations because the economic value of new technologies needs to be considered</li> <li>– Embrace and champion innovation</li> </ul> <p><b>Recommendations</b></p> <hr/> <ul style="list-style-type: none"> <li>– Facilitate broad, collaborative conversations that respect diversity</li> <li>– Frame national and global conversations on regulating AI in a coherent and accessible manner</li> <li>– Look at tangible and practical ways to show how regulation would work in practice and the potential interventions.</li> <li>– Develop new tools or approaches to build public confidence in AI and encourage investment and innovation</li> </ul>	<p><b>Focus</b></p> <hr/> <p>“How do we achieve transparency and build trust with people in AI, which is complex and constantly evolving?”</p> <p><b>Key takeaways</b></p> <hr/> <ul style="list-style-type: none"> <li>– Build trust by ensuring that human rights, including privacy, informs the full lifecycle of AI system from design through implementation and deployment</li> <li>– Ensure outcomes are people-centred</li> <li>– Prioritize accountability, fairness, safety and accessibility</li> <li>– Use willingness from around the world to work together to act as advisers and critical friends and support one another</li> <li>– Gain social licence for the use of AI by earning trust through transparency</li> </ul> <p><b>Recommendations</b></p> <hr/> <ul style="list-style-type: none"> <li>– Build social licence to use AI by designing national conversations on AI and its role in achieving national priorities and outcomes</li> <li>– Identify and iterate innovative approaches to assess algorithm and AI use, and conduct robust benefit and risk analyses</li> <li>– Develop regulatory capabilities through the design of an entity to oversee AI use</li> <li>– The three areas were refined and endorsed.</li> <li>– Areas worth further study, including possible pilots, were identified. There was agreement that the focus of the three areas provided a government with the right policy levers to influence the AI debate and gain traction.</li> </ul>

2

## Reimagining AI regulation

This project focuses on identifying and iterating innovative approaches and tools for regulating AI that can be scaled





“ The point of regulation is to provide certainty and to manage risks while allowing the benefits to be distributed equitably.

This section summarizes the key issues relating to the potential regulation of AI. It draws on the conversations that took place within the project communities through the various in-person and

online workshops and meetings, and demonstrates the path taken to the choice of the three focus areas for the project.

## What is regulation?

One of the first problems encountered in this project was the differing definitions people had for regulation. For the purposes of the project, regulation has been defined as the set of formal and informal rules, norms and sanctions that work together to shape people's behaviour to achieve a policy objective or goal. The point of regulation is to provide certainty and to manage risks while allowing the benefits to be distributed equitably. Regulation can also be used to consider new technologies that will increase human well-being, environmental sustainability and business innovation.

A regulatory system can use a range of tools, from hard law responses that set specific controls (these could include statutory legislation and regulation, rules, fines and subsidies), to soft power options such as awareness and education, partnerships, networking, consultation and engagement. Within this spectrum sit levers or tools such as guidance, codes of practice, charters, capacity and capability building, procurement, self-regulation, standards, certification, co-regulation, licencing, monitoring and auditing and enforcement.<sup>20</sup>

Governments are aware of the limitations on traditional regulatory systems; further, modern regulatory domains are so complex and dynamic that they can often no longer be handled by the state alone. In a number of countries, work is

underway to futureproof legislation and regulatory systems, using new digital tools<sup>21</sup> or approaches. Within the Centre for the Fourth Industrial Revolution, there is a project dedicated to looking at agile approaches to regulation, realizing that the twin challenges of fast-paced innovation and convergence of technologies may cause problems for governments attempting to regulate.<sup>22</sup> Many countries are looking at approaches that will allow them to graft a more agile and collaborative approach onto the existing regulatory practice. This frequently means more use of the soft law options that sit on the regulatory spectrum.

AI regulatory initiatives will rely heavily on an interdisciplinary and multistakeholder approach to address problems. Alongside this, the adoption of a more soft law approach will rely on greater industry accountability, with governments potentially sharing the burden of technology governance.

It is also worth noting that the question of AI regulation is intermingled with a range of other policy questions around emerging tech, including data protection, competition and antitrust, the algorithmic amplification of online content, and the geopolitics of tech. While these issues are related, there is analytical value to excluding them from the scope of the project as they have distinct diagnoses and potential solutions.

## An overview of the global focus on AI regulation

Many jurisdictions and bodies around the world are considering regulation of AI. In this section, the key policy developments in a few major areas are briefly reviewed. The substantial questions around AI that have prompted the consideration of AI regulation will be reviewed in detail in a later section.

In May 2019, the Organisation for Economic Co-operation and Development (OECD) adopted a set of intergovernmental policy guidelines on AI – building on prior work, including the IEEE's Ethically Aligned Design<sup>23</sup> – which have now been adopted by over 42 countries.<sup>24</sup> The guidelines aim to provide international standards that will ensure AI systems are designed to be robust, safe, fair and trustworthy. The same year, the Government of the United Arab Emirates released a set of ethical principles<sup>25</sup> aimed at encouraging organizations that deliver AI services to place a priority on fairness, transparency and accountability.

In Europe, European Commission (EC) President Ursula von der Leyden promised to regulate artificial intelligence as part of her public platform. In mid-February 2020, the EC began that process with the launch of a white paper<sup>26</sup> laying out its approach on AI, accompanied by a report on the safety and liability implications of AI.<sup>27</sup> The Commission also opened a public consultation process, framed around the white paper. Central to the EC strategy are three pillars: being ahead of technological developments and encouraging uptake; preparing for socio-economic changes; and ensuring appropriate ethical and legal framework.<sup>28</sup>

The United States has several initiatives in progress at the federal level. In early April 2019, the Algorithmic Accountability Act<sup>29</sup> was introduced to the US House of Representatives. If passed, the Act would require companies and organizations to conduct impact assessments to protect

against decisions that might otherwise be biased, inaccurate or unfair. Regulatory power would sit with the US Federal Trade Commission, which also looks after consumer protection and antitrust regulation. This Bill is still in Congress.

The current US administration has not been active in proposing regulation, aligning to general Republican Party policies. In January 2020, the White House issued a memorandum outlining a set of principles that are designed to promote innovation and growth. The memo urges regulators to make sure values such as transparency, risk management, fairness and non-discrimination are embedded into regulatory design. The Department of Defense also published a commitment to five high-level ethical principles – responsible, equitable, traceable, reliable, and governable – for AI use within the Department. Other initiatives have also taken place at the state and city levels, in New York, Illinois, California and elsewhere.<sup>30</sup>

**Several countries have followed the OECD's lead in developing a set of ethical principles that will underpin regulation. For example:**

- India has been working on processes and tools for the formulation of laws, guidelines and policies for governing and regulating AI, including the June 2018 NITI Aayog National Strategy for AI,<sup>31</sup> which recommended building a data privacy network to protect human rights and create regulatory guidelines, and setting up a task force to examine and issue modifications to existing laws.

- Canada's Office of the Privacy Commissioner launched a consultation process on appropriate AI regulation proposals in January 2020.<sup>32</sup> The 11 proposals set out for consideration in the discussion paper look at how privacy and rights issues, particularly in relation to data, should apply to developing and implementing AI and the future development of AI regulation.

- Japan published its draft AI R&D guidelines for international discussions<sup>33</sup> in July 2017, setting out for discussion non-binding AI R&D principles and guidelines to promote the benefits and reduce the risks of AI. Underpinning the guidelines was the importance of a human-centred society, sharing guidelines as soft law; ensuring the right balance of risks and benefits, not hindering the technologies or imposing excessive burdens on developers.

- In March 2019, Denmark launched its new National Strategy for Artificial Intelligence,<sup>34</sup> laying out four goals, the first of which is to ensure Denmark has a common ethical and human-centred base for AI use.

- In 2018, the Australia Federal Government pledged investment for improving Australia's capability in AI.<sup>35</sup> Part of this investment was given to develop an AI technology roadmap and an AI ethics framework.<sup>36</sup> The work closely follows the EU approach for AI regulation.

## Does AI need some form of regulation?

“ Existing regulatory regimes do not provide solutions for all the harms and issues, and existing regimes will not be fit for purpose in the future.

AI is not a wholly new and ungoverned space. As well as the work mentioned above, existing regulation in many countries already provides some protections, and countries can work within their existing legal and regulatory systems (for example, in New Zealand, some 80% of AI issues are covered by existing legislation). While this solution will work for the short term, existing regulatory regimes do not provide solutions for all the harms and issues, and existing regimes will not be fit for purpose in the future.

AI regulation is complex. There are no easy answers to the question of how best to regulate AI, including whether AI is regulated specifically or whether elements of it are regulated in different ways. The risks are known, however. Insufficient regulation creates uncertainty among AI developers and users and can result in areas where there are no legal protections, possibly creating more inequities and discrimination. This could result in social and economic failures, requiring more interventions to fix. Overregulation might lead to inhibition and stagnation, damaging innovation and global interoperability. At the moment, it is arguably the worst of all worlds – lack of transparency means

people distrust and are wary of AI and what they believe is happening, while in the absence of a clear authorizing framework, there is little willingness to innovate either.

As discussed previously, there are tensions in the possible regulation of AI. On the one hand, there is ample evidence that the use of AI will have wide-ranging impacts on citizens – it has the potential to provide many benefits, but it also carries the possibility of huge risk, including exacerbating existing divides and biases in our system. On the other hand, the full extent of these challenges and opportunities is not yet known. It is also unclear how AI will impact existing structures in the world, so expecting policy-makers to rush into designing suitable regulatory frameworks carries the risk of unenforceable regulation being written.

There are differing views from experts, governments and the private sector on what areas and activities should be regulated and approaches to regulating AI diverge sharply across regions. Emerging controversies surrounding AI can force the implementation of hastily constructed and suboptimal regulatory policies. Other projects

“ A gradual and iterative approach is required to reimagine regulation for the age of AI in a way that builds trust.

within the Centre for the Fourth Industrial Revolution portfolio are already tackling some of the riskiest uses of AI – facial recognition, human resources, and for applications targeting children, including education.

AI is not static. Both the range of uses and the technology itself are in a state of rapid development, which makes using traditional forms of regulation unworkable. AI crosses sector boundaries, and it operates across national boundaries, within new power structures that cannot easily be monitored, governed, or compliance enforced. At its heart, this issue is about the speed of regulatory change.

Further, a lack of consensus on a path forward and the risk of stifling innovation may deter any action. Before deciding to regulate, policy-makers try to determine what direct or indirect harms will come from AI, and whether they rise to the level where regulatory standards are needed.<sup>37</sup> This may be difficult as it is rare to have definitive answers when dealing with emerging technologies and policy-makers find themselves in the realm of

## Choice of project focus areas

Baker McKenzie has been tracking initiatives in the regulation of artificial intelligence globally for several years.<sup>40</sup> In work presented during the San Francisco workshop, Danielle Bennecke, Senior Associate at Baker McKenzie, identified common themes in regulatory proposals around AI across geographies as being accountability, fairness, human oversight, privacy, safety/security and transparency. The weight given to these themes differs across countries but together, these are the principles that underpin much of the design of regulation that is happening today.

**Rather than engage in a separate discussion on each one of the long list of issues, the project focused on the role of transparency as a fundamental enabler of high-quality AI governance. Questions explored in the workshops included:**

- What level of transparency is required for governments to gain social licence when using AI?
- What transparency requirements are needed to design risk assessment frameworks?
- How should the need for transparency shape the design of regulatory bodies?

These conversations highlighted more nuanced issues, such as how much information needs to be given to people – how can you make AI understandable to the ordinary person, and how much information do they really want? They have the right to explanation (e.g. the EU’s General Data Protection Regulation), but it is complicated and technical and often developers themselves

predictions and/or informed guesses. This decision-making needs to be done in consultation with, and based on direct input from, members of affected communities. This, in turn, requires that the public be informed about the existence of AI-based systems, or about the plans to develop and deploy such systems.<sup>38</sup> Once this decision is made, a more formal process is needed to define scope and understand context, so that the regulation provides the outcomes considered necessary. None of these things are easy with AI.

Regulating any kind of new technology is difficult. In addition to general issues relating to technology, regulating for AI is difficult from a regulatory design perspective. As mentioned above, this is not an unregulated space and, in many countries, existing laws are in place that cover technology and its impacts.<sup>39</sup> However, there are gaps (both of rules and of oversight by regulatory bodies), and increasingly, areas of concern where law does not exist, or is not fit for purpose to deal with the many complexities of AI and emerging technologies.

do not understand the AI and what happens in the black box. With that context, how do you provide transparency and build trust? There are also trade-offs with transparency such as IP for businesses and national security for governments. Governments and organizations need to explain the risks and how to mitigate the risks, because transparency can backfire if it is not thought through and implemented carefully.

Given these issues, a consensus was reached that a gradual and iterative approach is required to reimagine regulation for the age of AI in a way that builds trust. Co-designing the building blocks of a reimagined approach with key stakeholders and through multidisciplinary teams would allow for learnings to be incorporated into the approach, and for the scope and path to be refined, based on these learnings. The project team settled on three key areas of focus, to act as the building blocks.

The first area of focus will be developing a framework to run a national, widespread conversation with all stakeholders, seeking social licence and building trust in the process. A clear focus of this conversation will be to gain the input of those whose voices are often not heard, and who are likely to be impacted by the future use of AI systems.

Secondly, a number of conversations with organisations and government agencies showed that people would find it useful to have an independent group or body that could provide support and advice to users of AI. This body should not be a regulatory body, but should have a friendly face and set of functions that aimed at helping and supporting, not monitoring or

enforcing. The final stream of work is to explore what a Centre of Excellence for AI might look like.

Finally, trust is strengthened by having robust risk/benefit assessment built into the whole AI design and implementation process to mitigate their potential adverse effects. This led to the second stream of work, which is to build an assessment framework that is flexible enough for organisations

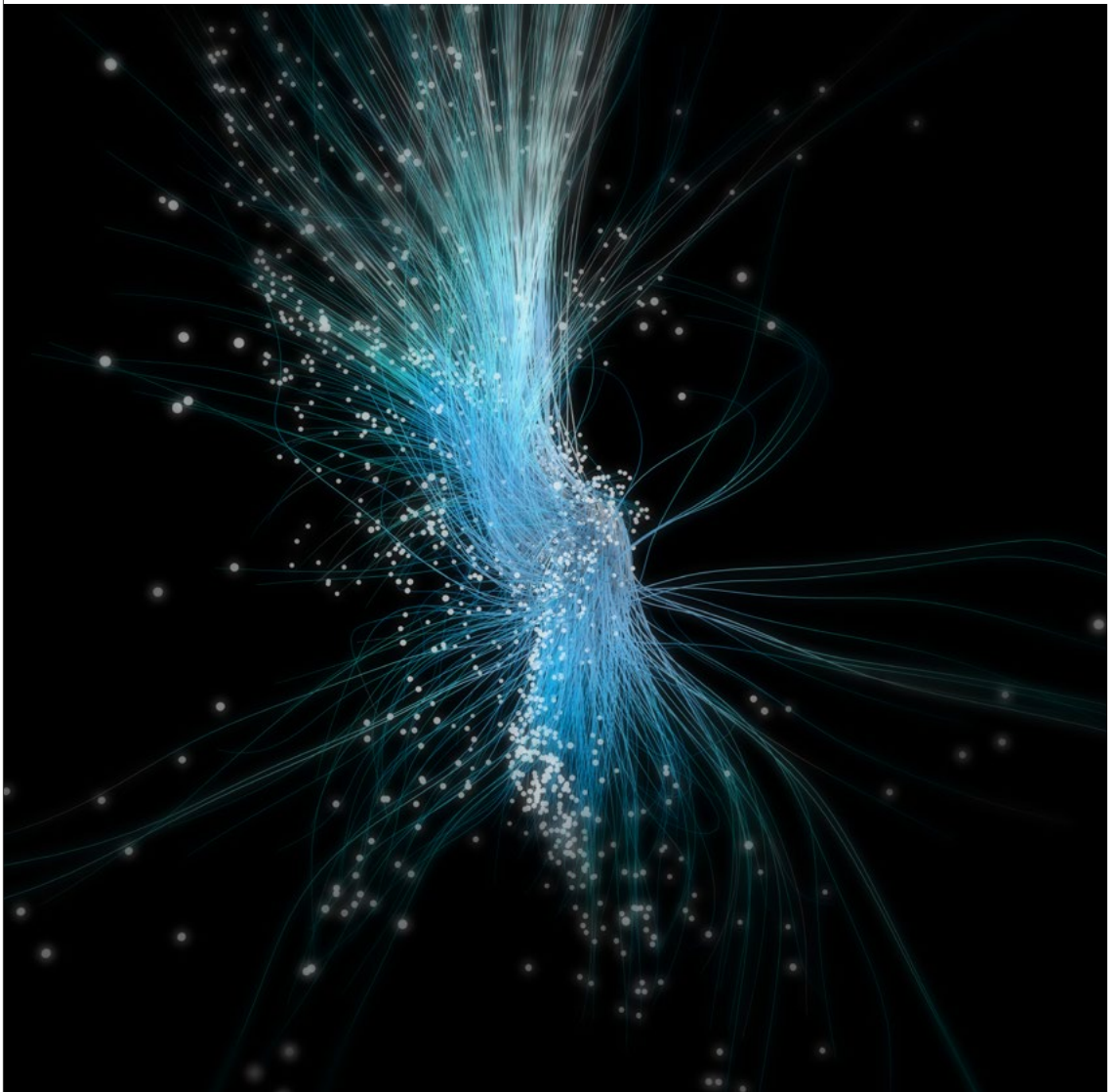
and governments to adapt it to their own particular situations.

Together, these three streams of work aim to build trust and transparency into the regulatory process. Used in conjunction with the other tools in the regulator's toolkit, these approaches and frameworks will provide useful steps in collaborative and agile regulation.

3

## Focus areas

The project community has identified three focus areas and co-designed actionable frameworks within each of them



## 3.1 National conversation

### An important tool to gain social licence

Trust is crucial for acceptance of AI and gaining the social licence to use its full potential. This trust is built through open and honest conversations with the people who will be affected by AI. This is not just an issue for governments and their citizens; businesses also have an interest in hearing from their customers. Thus, all regulatory conversations need to start with a more basic set of conversations with your community – Who are we? What do we value? How are decisions made in society? How is power distributed? Who will benefit? Who will be excluded or harmed? And are we happy with this status quo?

“ Social licence is about communities agreeing that governments, agencies, or companies are considered trustworthy enough to undertake activities that may have risks.

As discussed earlier, social licence is about communities agreeing that governments, agencies, or companies are considered trustworthy enough to undertake activities that may have risks. National conversations are an important tool to gain this social licence and trust. They allow for the tailoring of messages to a wide variety of people across society, giving them the tools and platforms to learn about issues, ask questions and help co-design solutions, all within an equitable and respectful environment. Building an environment of trust and inclusion will also encourage more people to participate, creating more equity and serving the interests of more and more groups and communities.<sup>41</sup>

#### There are already a number of examples of good practice in social licence:

- The New Zealand Ministry of Social Development has its Privacy Human Rights and Ethics (PHRaE) framework,<sup>42</sup> which informs decisions it makes about initiatives and which has been extensively tested with its stakeholders; and the Social Wellbeing Agency (formerly the Social Investment Agency), which has undertaken thorough consultation on the ethical principles it uses to inform its work.<sup>43</sup>
- Also in New Zealand, the Data Futures Partnership framed social licence in 2017 as the acceptance by individuals for organizations to use their data, information and stories. They identified value, choice and protection as critical factors associated with trusted use of personal data and information.<sup>44</sup>
- The Australian Human Rights Commission is currently leading a project on Human Rights and Technology.<sup>45</sup> The aim is to advance human rights protection during vast technological change and look at how levers and measures (such as law, policy and incentives) can promote and protect human rights in a technological age. The project is seeking views from across Australian society,

using a range of engagement tools and approaches, and will present options informed by these views to the federal government.

For governments, getting this public agreement is a key part of democracy, whereas for businesses, it is central to their success that their customers trust them and the services they provide. The two workshops with the global community identified a set of issues and success factors for national conversations. As mentioned earlier, trust is the single most important factor for success. The process of engagement must be undertaken in good faith and people know that their input will be valued. Equally important is trustworthiness, with evidence being used to show people that the thing they are being consulted on is trustworthy.

#### Trust can be enhanced by some the below factors (this is not an exhaustive list):

- **Veracity of information:** Trust can be lost if people feel manipulated, with facts and data that seem biased or selective, or that clearly reflect vested interests or the views of powerful players in the process.
- **Independent expert:** Having a trusted third-party to broker agreement between the public, politicians and technical experts. They can translate data and provide information that is reliable, accurate and easily understood and to work as the go-between in building trust.
- **Ability to influence:** There should be clarity around the participants' ability to influence decisions. To maintain trust, the organization undertaking the consultation needs to be open and honest about the level of engagement being sought and what weight will be given to their views.
- **Seeking a broad set of views:** To be successful, conversations will seek out and include views which reflect the diversity and spread of the community being consulted. Consultation should also reflect the cultural needs of the group whose views are being sought.
- **Hearing all voices:** Strong voices can predominate in conversations. Sometimes these are lobby groups, reflecting minority views, and other times the majority view overwhelms the concerns and specific views of minority groups. The right opportunities need to be provided to allow all interested or affected parties to have their say.

- **Education and awareness-raising:** Proper information is vital in this process. It takes time and energy to understand complicated issues or navigate bureaucratic processes so people should be empowered, educated, and enabled to make decisions on technology ethics and rights before they can engage.

Badly done collaboration and engagement will create resentment and distrust, so several practical matters need to be considered by the designers for a successful national conversation. There are a large number of toolkits that already exist to help guide engagements. Part of the next phase of work will be to compile a list of these tools.

As a final note, the team recognizes that national priorities will inform the national conversation work, and to be successful, it is useful to nest domestic AI

work in an existing narrative frame that already has buy-in from political leaders and key stakeholders. This can complement other work such as national strategies that are underway.

Regulating AI well, however, means striking the right balance between rules or regulations that have a global applicability and those that are relevant to the local context. Domestic conversations will give an understanding of where people want to sit on the spectrum of AI use in their communities but, as part of a global community, it is also important to identify global views and expectations of AI use. There may be many similarities, which gives useful data to nations looking to have a global approach to AI operability. Where there are differences between groups and nations, there may be some value in looking for new ways to work together to co-design approaches.

## A starting point

The below steps are suggested as a starting point to facilitate national discussions on AI. The process is written at a general level, so it could be used both for nation/state wide discussions on high-level issues or by an agency or organization on a specific tool where community/public agreement

to proceed is being sought. This is not a linear process. Steps may merge together or happen more naturally at different parts, or may need to be repeated throughout the process, or could be missed out, depending on the engagement scope and outcomes.

## Pilot project:





### Refining tools for organizations willing to have such conversations

Through this work, a six-step plan for holding national conversations has been developed. Over the next phase, the project team will use the plan to help stakeholders hold conversations with stakeholders on AI issues. In particular, the team wishes to see more engagement with those who traditionally have had less voice in policy decisions. By involving all strata of society, a consensus can be built on AI ethics and values that will underpin AI

policy and use. This will allow people from all parts of the community to be actively involved in choices regarding AI use.

The team will also work with those who have conducted successful large-scale engagements across the world to produce a set of case studies that identify key success factors, and will build a repository of toolkits already in use.

FIGURE 2 | Steps for holding a national conversation

<p>Define</p> 	<p>Develop a set of principles and values that reflect the societal context of the government/ organization. These principles and values will underpin and inform the conversation (there are already many AI principles in existence that may help with this process).</p> <p>In some cases, consultation on the principles and values may be one of the first things consulted on.</p>	<p>In other cases, projects will want to demonstrate to their audiences that their work rests on a strong ethical base.</p> <p>This stage will also include contextual information on areas such as privacy, security and user control. These issues will be of concern to people, and different countries have different legislative frameworks to manage these issues.</p>
<p>Discovery</p> 	<p>Know the current context and how the engagement is intending to get to a new, desired state. This can include the wider context of a state's existing regulatory and technology system, or be the context within which a specific AI system sits.</p> <p>A discovery stage is undertaken to develop an understanding of the issue, the rationale for engagement, scope of the engagement and goals or outcomes to be achieved.</p>	<p>It should also be clear on who the audience is for the results of the engagement, and any limitations for the work.</p> <p>For specific consultations, this stage will also help designers decide whether AI is right for their particular use case.</p>
<p>Decide</p> 	<p>This step allows planners to make key decisions on the shape of the engagement.</p> <p>It would be useful to use a social intelligence approach to understand the social dynamics of the society/groups being consulted, and to identify the best approaches to reach those audiences.</p> <p><b>Central issues to address could be the:</b></p> <ul style="list-style-type: none"> <li>– Make-up of the participants, ensuring diversity and representation from all sectors of society</li> </ul>	<ul style="list-style-type: none"> <li>– Level of engagement to be undertaken and the impact that the views will have on actual decision-making (for example, people may be consulted, informed, asked to participate or collaborate, etc.)</li> <li>– Channels and tools to be used for communication.</li> </ul> <p>For more targeted engagements on specific AI models or tools, this step is a chance to examine the training data to ensure the data held is representative and unbiased.</p>
<p>Design</p> 	<p>Designing the actual engagement, utilizing values-driven, human-centric methodologies for every step of the value chain.</p> <p>Choices here will be informed by the intended participant list, level of involvement sought by the participants, and the nature of the feedback being sought.</p>	<p>There are a wide range of tools available and an engagement strategy would be useful.</p> <p>Also important at this step is an understanding of what educational or awareness-raising material might be necessary for participants.</p>
<p>Analyse</p> 	<p>Once the engagement has occurred and information gathered, the results need to be analysed and synthesised. Work will need to occur on looking at the inputs and adjusting for bias, vested interests, monopoly of views, strong voices.</p>	<p>The findings will need to be presented to key stakeholders and to other key audiences, including the participants. These findings will include suggestions on how to use the research and next steps.</p>
<p>Review</p> 	<p>The final step is an evaluation of the process. It should involve follow-up with the participants to get their views.</p>	<p>Ideally, it would also provide an evaluation of how the work was used in decisions and informed policy-making.</p>



## 3.2 Regulatory capabilities and institutional design

The project community has considered a range of different regulatory capabilities required and options for the design of institutions to organize those capabilities. For example, a new institution or regulatory agency could:

- Have responsibility for a strategy or framework that informs the development of AI policy
- Be a single standards setter
- Be a best practices sharing group
- Have powers to encourage or incentivise the ethical use of AI
- Be the place that identifies gaps in existing law and suggests solutions

- Be a development group that collaborates to create something, or lead on short-term projects or milestones
- Consult with and advise existing regulators<sup>46</sup>
- Identify and advise on common, global and futures challenges and impacts<sup>47,48</sup>

The project is currently focusing on two main options when it comes to regulatory capabilities: a new regulatory body to oversee AI use, or a centre of excellence for AI capacity across government. This section explores both of these ideas in turn before proposing concrete pilot ideas.

### A new regulatory body to oversee AI

A number of recent proposals suggest a specific body to regulate the use of AI and algorithms – either in government specifically or its general use across the economy. Regulatory bodies can have a range of forms, functions and powers. In fact, a regulatory body offers a number of advantages over the regulation via legal and legislative systems.<sup>49</sup> They can be tailored for the regulation of a particular industry, or to address a specific problem. They can be staffed by experts rather than the generalists, and they can undertake investigations and make broad policy decisions, rather than reacting to legal cases brought.

As previously discussed, regulation is more than just binding rules imposed by a government. Modern regulatory domains are so complex that they can no longer be handled by the state alone. A modern regulatory body is not authoritarian, but one part of a broader system that influences and steers through a full spectrum of approaches. Harder regulatory powers may be useful in certain situations, while soft approaches will work better in others.

Participants at the San Francisco workshop raised objections to proposals for a new regulatory body, reflecting and extending criticisms present in the literature. They drew attention to the problem of defining the field of operation and scope of competence given the breadth of uses of AI. This relates to a second critique – the lack of tractability of putting all algorithms into one bucket without proper analysis of the level of risk and need for oversight. Both these points were visible in widespread commentary in the 2019 German Data Ethics Commission proposal for a horizontal classification and regulation of all algorithms.

Domain-specific regulatory responses may be required, reflecting both the daunting need for domain-specific knowledge (either technical AI expertise or context-specific knowledge in the area of the AI application in question) and the extent to which existing regulations and regulatory bodies already cover some aspect of AI systems. The Otago University report<sup>50</sup> on AI in government recommends that individual bodies have oversight of AI devices intended for use in a particular context. Sector or domain-specific regulatory approaches to AI could then be connected in a network of expertise.

Some of these challenges have obvious remedies. For example, a lack of experience can be solved through giving the group the ability to call on expert advice when it is needed. Other challenges are more complex. While regulatory scope can be defined based on the context, too narrow or specific a scope may work against regulators in areas of fast-moving technology. Some regulators (such as the US Food and Drug Administration and New Zealand's proposed therapeutic products regulator) have deliberately broad remits that are intended to future-proof it if the field moves in unforeseen directions.

As discussed previously, the development and use of AI is driven by global technology companies and thus a national regulatory body has limitations. An international AI regulatory agency could create a unified framework for the regulation of AI, inform the development of AI policies around the world and develop new norms around the use of AI.<sup>51</sup>

“ A key starting point for the work would be to map the landscape of existing regulation of AI and its use across government to identify gaps and overlaps.

## A centre of excellence for AI

The idea of a centre of excellence emerged within the project community as the group discussed the best way for individual governments to shape the development of ethical AI practices both within government and in their country. The United Kingdom was recognized as a pioneer in this area both through its Office for Artificial Intelligence and the Centre for Data Ethics. The Canadian experience in creating algorithmic assessment requirements across government was also cited, as was the Danish Expert Group on Data Ethics.<sup>52</sup> Further developing a structured experiment with a centre of excellence for AI across government (but also involving the private sector) – in New Zealand or elsewhere – would represent a significant contribution to the global conversation on the beneficial use of AI.

Such a centre should adopt a multidisciplinary and collaborative approach, with staff drawn from a range of disciplines, including civil society, academia, government and a wide range of industry players. It could be innovative and fluid, changing out staff depending on the issues being looked at.

While it should work closely with all the interested stakeholders, it will need to be sufficiently independent from government and industry to maintain credibility. For trust in it to be built and maintained, it cannot be limited in its assessments and recommendations by what is palatable to stakeholder interests. A key starting point for the work would be to map the landscape of existing regulation of AI and its use across government to identify gaps and overlaps and identify where a regulatory agency or a centre of excellence could add value.

### A range of functions was suggested for the centre of excellence, including:

- Gathering information and intelligence on AI use and providing reporting and reports

## New Zealand context

The *Government Use of Artificial Intelligence in New Zealand* report recommends the New Zealand government consider the establishment of an independent regulatory/oversight agency that could:

- Work with individual government agencies looking to introduce new predictive algorithms or use an already deployed one for a new purpose
- Maintain a register of uses of predictive algorithms within government agencies

- Ombudsman-type role in transparently highlighting problems in the use of algorithms and AI
- Error reporting (e.g. for commercial products where it can help with product safety, allowing for users to update or change algorithm use where problems have been identified, and investigating how errors affect different cohorts)
- Identifying of risks and solutions in algorithms
- Seeking citizen-centred solutions through the use of human-centred design approaches and skills
- Providing advice and support (e.g. to the government, private sector, individual organizations, civil society, designers of AI, procurers of AI, etc.)
- Work with individual agencies and government ministries to assist in their development and procurement of algorithms
- Pooling expertise and best practices on the development and use of AI solutions that can be scaled across government and the private sector
- Developing regulatory sandboxes
- Principles and principle-based approaches that reflect the diversity of communities, including indigenous views
- Applying well-being and sustainability metrics to the evaluation of government algorithms
- Identifying key skills and competencies needed for AI positions
- Identifying, creating and disseminating information on educational and workforce development opportunities for employees on AI topics

- Receive, at specified intervals, ongoing assessments of the use of algorithms, from individual agencies
- Oversee the use of “self -checking” frameworks

Establishing such an agency would involve consideration of a number of questions, including the precise extent of its remit and making sure there is no unnecessary duplication of effort or overlapping remit with existing agencies, such as the Office of the Privacy Commissioner.

## Pilot project:

### Partnering with governments seeking to strengthen their regulatory capabilities

Useful pilots could be designed by partner governments working with the Centre for the Fourth Industrial Revolution to explore either the new regulatory agency concept, or the centre of excellence, or both. Mapping the existing regulatory landscape will be a crucial step to discerning the right pilot for any country. In general, less well-developed regulatory environments may see greater benefit from a new body. Furthermore, given the design of a new institution should reflect the

concerns and issues of the nation it serves, such a pilot would connect well with a parallel pilot in a national conversation.

In New Zealand, initial pilot planning aims to set up a small working group of interested experts to develop a discussion and options paper. The options paper could then be presented to the New Zealand government to consider the potential establishment of a new agency or centre of excellence.

FIGURE 3 | Key questions to explore in the design of options

Case for change	<ul style="list-style-type: none"><li>What is the context and landscape?</li><li>What is the rationale for deciding to create an agency?</li><li>What problem is being addressed?</li></ul>
Vision	<ul style="list-style-type: none"><li>What is the future state you want?</li><li>How will you tell this story to your audience?</li><li>What outcomes are you looking to achieve from this agency being set up?</li></ul>
Mandate and power	<ul style="list-style-type: none"><li>Where does the mandate for the proposed centre of excellence come from, i.e. who is the sponsor for this options analysis and, once up and running, where would the centre draw its authority from?</li><li>What powers does the agency need to discharge these functions? What is the nature and scope of these powers?</li><li>Will it involve the public and private sectors? If so, how will it approach the differing legal obligations and behavioural drivers of each?</li></ul>
Functions and form	<ul style="list-style-type: none"><li>What is the role of the agency to be? Where in the current system will it sit? What are its key linkages?</li><li>What activities will the agency do initially? How will that evolve over time?</li><li>What form will this new agency take?</li></ul>
Resources	<ul style="list-style-type: none"><li>What funding and resource is needed to run the new model?</li><li>Where will this resource come from?</li><li>Does it have access to the right expertise? Can it access external expertise?</li></ul>
Governance and accountability	<ul style="list-style-type: none"><li>What sort of leadership does this agency need?</li><li>How will the centre be held accountable for its work?</li><li>It might be required to report publicly on a summary of its activities, including specific projects worked on. The report could make recommendations on ways to better support the development and deployment of AI.</li><li>How does the agency cooperate with international partners to meet AI's potential and anticipated impact on all of humanity?</li></ul>

## 3.3 Risk/benefit assessment of AI systems for government

### The need for increased oversight of AI systems used by government

The project community strongly believes that developing internal regulatory capabilities for governments is an essential step towards the trustworthy use of AI. However, this is not enough to achieve the overarching goal: to operationalize the ethical use of AI (i.e. ensuring that AI systems operate in accordance with laws and social norms). This is particularly important for systems used by government agencies where AI-powered services hold the potential to vastly improve their operations and help meet the needs of citizens in new ways, ranging from healthcare delivery<sup>53</sup> to personalized learning.<sup>54</sup>

As explained in previous sections, government officials are increasingly aware of the transformational impact of artificial intelligence. Yet, they have not fully embraced this technology for the public sector because they know that without proper oversight, AI may be ineffective or, worse, lead to unethical outcomes, erosions of individual privacy and security, and abuses of human rights. They are also uncertain about the oversight process that should be introduced to ensure that AI systems do not violate a society's norms for due process, which generally guarantee that all legal proceedings will be fair and that one will be given notice of the proceedings and an opportunity to be heard, by a human judge (or panel of judges), before the government acts to take away one's life, liberty, or property.<sup>55</sup>

A failure to guard against such violations can lead to the erosion of fundamental assumptions of a society's judicial system and the undermining of the rule of law itself. Such processes are therefore required to address legitimate technocratic and democratic concerns. The former refers to the need for government agencies to inform themselves sufficiently about the functioning of their AI systems (training set, optimization function, limitations of the model selected, etc.).<sup>56</sup> The latter is related to the importance of reporting publicly the decisions that have been made by AI systems, primarily to inform citizens affected by them.

To address these challenges and unlock the benefits of AI for the public sector, the project community calls for the introduction of risk/

benefit assessment frameworks for AI systems used in government. They perceive this as a key complement to the development of in-house expertise on AI.

Governments are already designing risk/benefit assessment frameworks. In 2018, the Government of Canada presented its Algorithmic Impact Assessment<sup>57</sup> (AIA), a questionnaire designed to help assess and mitigate the risks associated with deploying an automated decision system within a government agency. In 2019, the UK government released its new procurement guidelines<sup>58</sup> for AI-powered services to inform and empower buyers in the public sector, helping them to evaluate suppliers, then confidently and responsibly procure AI technologies for the benefit of citizens. In 2020, Singapore released the second edition of its Model AI Governance Framework,<sup>59</sup> which converts relevant ethical principles to implementable practices in an AI deployment process.

These frameworks have been developed and adopted to meet multiple objectives. First, the frameworks are intended to enable governments to fulfil their duty to inform their citizens about the AI systems that affect their lives by documenting the design of these systems, their purpose and the context of use. Second, these frameworks have been designed with the objective of ensuring the effective identification, monitoring and mitigation of the risks associated with specific AI systems by solidly grounding AI systems with assessment criteria and usage scenarios. Third, they are intended to provide greater accountability by providing a meaningful and ongoing opportunity for external review of AI systems. Indeed, any appointed competent third party should be able to evaluate whether a specific AI system is trustworthy by examining its associated risk/benefit assessment framework. Fourth, such frameworks should increase public agencies' internal expertise on the AI systems that they deploy, building organizational capability in this domain. Finally, a well-designed framework should create the space for contestability – any citizen should be able to challenge the decision made by a specific AI system by using such documentation.

“ These frameworks have been designed with the objective of ensuring the effective identification, monitoring, and mitigation of the risks associated with specific AI systems.

## New Zealand context

The Ministry for Social Development of New Zealand has taken a leading role in the effort to ensure that risks that the government could be impinging on people's privacy, human rights, or ethics when designing a new service using personal information are identified early on. Using people's data is central to its work, and it has made a commitment to respect individual privacy and to be clear about how it uses and shares information. To this end, the ministry worked with an independent university ethicist in 2017 to develop a tool that would improve its approach to responsibly using and safeguarding personal information. This tool is named the Privacy, Human Rights and Ethics framework (PHRaE).<sup>60</sup>

PHRaE is a risk assessment tool that civil servants from the Ministry of Social Development must use when they design any new service, including AI-powered services and automated decision systems, to ensure the identification and mitigation of risks at the beginning of the design process. This enables PHRaE risks to be designed out rather than risk being a barrier to implementation. It is supported by a team of specialists that engages with project teams in an iterative and active discussion throughout the project's life cycle and explicitly takes a user-centric approach. In this

regard, it significantly departs from the principle-based risk assessment framework designed by some governments.

The tool provides various advantages. First, it encourages project teams to actively consider the potential risks and benefits associated with their new services and what assumptions would need to be true to deliver the perceived added value. Second, it enables them to exclude unnecessarily harmful projects and introduce risk mitigation strategies when needed. Third, it leads to the creation of a repository of the services envisioned – those that got rejected and the ones that got through. Over time and through multiple projects, the ministry will build strong organizational capability in the responsible use of personal information.

PHRaE is still a pilot project. As such, it is going through a thorough review process that includes feedback from project teams, and internal and external stakeholders. The Ministry for Social Development is currently developing a prototype of an online, interactive assessment tool via an iterative process to roll the pilot across departments. In this spirit, it would be pleased to collaborate with governments that are developing similar tools to strengthen their current prototype.

### Pilot project:

#### Guidelines to ensure the responsible development of risk/benefit assessment framework for AI in government

During the pilot phase, the aim is to partner with governments that are yet to develop their frameworks, or which are in the process of testing their existing frameworks, starting with New Zealand and their PHRaE framework. To this end, the project community has identified a set of guidelines that will help them to ask the right questions, follow the best practices, identify and involve the right stakeholders in the process, and ultimately create a sensible risk/benefit assessment framework.

These guidelines do not represent the final word on the risk/benefit assessment framework discussion. This would not be reasonable because determining a single comprehensive set of risk/

benefit assessment framework elements is likely infeasible as the context that project teams face will vary a lot across departments and governments. In fact, modulation in the application of the framework is required if it is to be effectively operationalized across a range of different AI-powered systems, each entailing different risks and different levels of risk. Rather, these guidelines represent the first step in an interactive conversation on the types of elements that should be included in such frameworks building on the existing literature.<sup>61</sup> The Centre for the Fourth Industrial Revolution welcomes stakeholders willing to take part in this discussion to join our project.

FIGURE 4 | Key considerations for the design of risk/benefit assessment framework

Guidelines	Rationale
<p><b>1. Justify the choice of introducing an AI-powered service</b></p>	<p>Before considering how to mitigate the risks associated with AI-powered services, governments willing to deploy them should clearly lay out their assigned objectives and how they are supposed to benefit various stakeholders (e.g. end users, consumers, citizens and society at large).</p>
<p><b>2. Adopt a multistakeholder approach</b></p>	<p>Project teams should identify the stakeholders across government, civil society, academia and the private sector that should be anchored to this particular project and provide them with relevant information about the usage scenarios envisioned and the specification of the AI system under consideration. Special attention should be paid to disadvantaged and marginalized communities and give them relevant educational material to enable their meaningful participation in the consultation process.</p>
<p><b>3. Consider relevant regulations and build on existing best practices</b></p>	<p>When considering the risks and benefits associated with specific AI-powered solutions, include relevant human and civil rights in impact assessments. It is also important to leverage the lessons learned from governments that have developed similar risk assessment frameworks.</p>
<p><b>4. Apply risks/benefits assessment frameworks across the lifecycle of AI-powered services</b></p>	<p>An important distinction between AI software and services from traditional software development is the learning aspect (i.e. the underlying model evolves with data and use). Therefore, any sensible risk assessment framework has to integrate both the build-time (design) and run-time (monitor and manage). Also, it should be amenable for assessment from a multistakeholder perspective both at build-time and run-time.</p>
<p><b>5. Adopt a user-centric and use case-based approach</b></p>	<p>To ensure that government risks/benefits assessment frameworks are effectively actionable they should be designed from the perspective of the project teams and around specific use cases.</p>
<p><b>6. Clearly lay out a risk prioritization scheme</b></p>	<p>Diverse groups of stakeholders have different risk/benefit perceptions and levels of tolerance. Therefore, it is essential to implement processes explaining how risks and benefits are prioritized and competing interests resolved.</p>

Guidelines	Rationale
<p><b>7. Define performance metrics</b></p>	<p>Project teams, in consultation with key stakeholders, should define clear metrics for assessing the AI-powered system's fitness for its intended purpose. Such metrics should cover the system's narrowly defined effectiveness or accuracy as well as other aspects of the system's more broadly defined fitness for purpose (including regulatory compliance, user experience, adoption rates, etc.).</p>
<p><b>8. Define operational roles</b></p>	<p>Project teams should clearly define the roles for humans in the deployment and operation of any AI-powered system. The definition should include clear specification of the responsibilities of each human agent required for the effective operation of the system, the competencies required for filling the role and the risks associated with a failure to fill the roles as intended.</p>
<p><b>9. Specify data requirements and flows</b></p>	<p>Project teams should specify the volumes and nature of data required for the effective training, testing and operation of any AI-powered system. Project teams should map data flows expected with the operation of the system (including data acquisition, processing, storage, and final disposition) and identify provisions to maintain data security and integrity at each stage in the data lifecycle.</p>
<p><b>10. Specify lines of accountability</b></p>	<p>Project teams should map lines of responsibility for outcomes (both intermediate and final) generated by any AI-powered system. Such a map should enable a third party to assess responsibility for any unexpected outcome of the system.</p>
<p><b>11. Support a culture of experimentation</b></p>	<p>Governments should advocate for a right to experiment around AI-powered services for deployment to encourage calculated risks. In practice, this requires setting up feasibility and validation studies, encouraging cross-collaboration across departments and fields of expertise, sharing of knowledge and feedback via a dedicated platform.</p>
<p><b>12. Create educational</b></p>	<p>Building a repository of various risks/benefits assessment frameworks, their performance and revised versions to develop strong organizational capability in the deployment of AI-powered services is key.</p>

# Conclusion

This white paper is the first step in an iterative process, and we welcome participation of organizations willing to engage in this topic

Artificial intelligence is impacting society at unprecedented speed, scope and scale. This creates unique opportunities and challenges. Maximizing the benefits of AI while mitigating its adverse effects requires significant adjustments to the existing regulatory environment, and policy-makers around the world are increasingly acknowledging this need. Yet, at times, the complexity of this task appears daunting. Should we adopt a transversal or a sectoral approach? Should we favour ex-ante or ex-post intervention? Do we need a regulatory agency for AI? When starting from a conceptual perspective, we are immediately faced with complex questions, whose answers seem to only lead to more questions.

We took a different approach where we primarily looked for tools that could be leveraged by national governments. To this end, we have built a multistakeholder, evidence-based policy project anchored in New Zealand. We think that our global community of experts has achieved sensible

progress in identifying some of the key focus areas where those adjustments are the most actionable: national conversations to gain a social license; regulatory capabilities; and risk/benefit assessment of AI systems within government.

For each of these areas, we have produced frameworks and guidelines that, once combined, contribute to the development of an appropriate regulatory environment for AI. In the next phase of this project, we will test these frameworks and guidelines, assess their relevance and review them based on the observed results. We will start in New Zealand and across various jurisdictions by leveraging our network government partners. Then, we will share the lessons learned.

Considering our open and experimental approach, we encourage government officials, industry players, civil society representatives and academics to join us on this journey to strengthen our frameworks and ensure their greater impact.



# Contributors

## Lead authors

### Lofred Madzou

Project Lead, AI and Machine Learning,  
World Economic Forum

### Michael Costigan

Salesforce Fellow,  
World Economic Forum

### Kate MacDonald

New Zealand Government Fellow,  
World Economic Forum

## Contributors

### Kay Firth-Butterfield

Head of Artificial Intelligence and Machine Learning,  
World Economic Forum

### Emma Naji

Executive Director, AI Forum, New Zealand

### Colin Gavaghan

Director, New Zealand Law Foundation Centre for Law  
and Policy in Emerging Technologies

### James Maclaurin

Associate Professor, Centre for Artificial Intelligence  
and Public Policy, University of Otago, New Zealand

### Alistair Knott

Associate Professor, Centre for Artificial Intelligence  
and Public Policy, University of Otago, New Zealand

### Joy Liddicoat

Assistant Research Fellow, Artificial Intelligence,  
University of Otago, New Zealand

### Olivia Erdelyi

Lecturer, Researcher on AI Governance, Canterbury  
University, New Zealand

### Iven Michiel Yvonne Mareels

Lab Director, IBM Research Australia, Australia

### Sara Cole Stratton

Ngati Hine, Ngati Kahu, New Zealand

### Fredrik Heintz

Professor of AI, Linköping University, Sweden

### Kelly Pendergrast

Founder and Co-Director Antistatic, New Zealand

### Anna Pendergrast

Co-Lead, Antistatic, New Zealand

### Toby Walsh

Professor of AI, University of New South Wales,  
Australia

### Ilana Golbin

Director, AI Accelerator, PwC, USA

### Kumar Bhaskaran

Programme Director, Impact Science, IBM Research,  
USA

### Julia Stoyanovich

Assistant Professor of Computer Science, Engineering  
and Data Science, New York University, USA

### Ashley Casovan

Executive Director, AI Global, Canada

### Bruce Hedin

Principal Scientist, H5, USA

### Kathy Baxter

Architect, Ethical AI Practice, Salesforce, USA

### Yoav Schlesinger

Principal, Ethical AI Practice, Salesforce, USA

### Daniel Lim

Salesforce Fellow, World Economic Forum

### Maria Luciana Axente

Lead for Responsible AI  
& AI for Good Lead at PwC UK

### Nils Gilman

Vice-President of Programs, Berggruen Institute, USA

### Danielle Benecke

Senior Associate, IPTech, Baker McKenzie, USA

### George Tilesch

Senior AI Strategist and Co-Author, Between Brains,  
USA

### Rumman Chowdhury

Managing Director and Global Responsible AI Lead,  
Accenture, Ireland

### Aaron Rieke

Managing Director, UpTurn, USA

### John Havens

Executive Director, Global Initiative on Ethics of  
Autonomous and Intelligent Systems, Institute of  
Electrical and Electronics Engineers (IEEE), USA

### Mikael Ekman

Chief Advisor and Deputy, Office of the Tech  
Ambassador, Denmark

**Mette Vestergaard Dam**

Tech and Cyber Advisor, Office of the Tech  
Ambassador, Denmark

**Horst Eidenmueller**

Freshfields Professor of Commercial Law,  
Faculty of Law, University of Oxford, United Kingdom

**William Eggers**

Executive Director, Deloitte Center for Government  
Insights, Deloitte, USA

**Ryan Clough**

Senior Director of Public Policy and General Counsel,  
Association of Research Libraries (ARL), USA

**Roger Brownsword**

Professor of Law, King's College London,  
United Kingdom

**Marietje Schaake**

Director, International Policy, Cyber Policy Center,  
Stanford University, USA

**Matthew Scherer**

Associate, Littler Mendelson, USA

**Clara Blume**

Deputy Director and Head of Art, Science,  
and Technology, Open Austria, USA

**Martin Rauchbauer**

Co-Director, Open Austria, USA

# Endnotes

1. "How AI is changing the customer experience", MIT Technology Review, 28 April 2020, <https://www.technologyreview.com/2020/04/28/1000675/how-ai-is-changing-the-customer-experience>.
2. "Artificial intelligence is the future of growth", Accenture, <https://www.accenture.com/sk-en/insight-artificial-intelligence-future-growth>.
3. Deloitte, Autonomous Driving: Moonshot Project with Quantum Leap from Hardware to Software & AI Focus, January 2019, [https://www2.deloitte.com/content/dam/Deloitte/be/Documents/Deloitte\\_Autonomous-Driving.pdf](https://www2.deloitte.com/content/dam/Deloitte/be/Documents/Deloitte_Autonomous-Driving.pdf).
4. World Economic Forum, Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data, October 2019, <https://www.weforum.org/whitepapers/federated-data-systems-balancing-innovation-and-trust-in-the-use-of-sensitive-data>.
5. Müller, Vincent C. and Nick Bostrom, Future Progress in Artificial Intelligence: A Survey of Expert Opinion, Springer, 2016, [https://link.springer.com/chapter/10.1007/978-3-319-26485-1\\_33](https://link.springer.com/chapter/10.1007/978-3-319-26485-1_33)
6. Barocas, Salon and Andrew D. Selbst, Big Data's Disparate Impact, SSRN, 2016, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2477899](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899).
7. Benedikt, Frey, Carl and Michael A. Osborne, The future of Employment: how susceptible are jobs to computerisation?, Oxford Martin School, 2013, [https://www.oxfordmartin.ox.ac.uk/downloads/academic/The\\_Future\\_of\\_Employment.pdf](https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf).
8. "Unintended Consequences of Biased Robotic and Artificial Intelligence Systems". IEEE Robotics & Automation Magazine, September 2019, <https://ieeexplore.ieee.org/document/8825881>.
9. Angwin, Julia and Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias", ProPublica, 23 May 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
10. "We designed an experimental AI tool to predict which COVID-19 patients are going to get the sickest", The Conversation, 14 May 2020, <https://theconversation.com/we-designed-an-experimental-ai-tool-to-predict-which-covid-19-patients-are-going-to-get-the-sickest-136125>.
11. Knight, Will, "The Apple Card Didn't 'See' Gender—and That's the Problem", Wired, 19 November 2019, <https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem>.
12. Yang, Jenny R. Ensuring a future that advances equity in algorithmic employment decisions, 5 February 2020, statement made to the Civil Rights and Human Services Subcommittee, Committee on Education and Labor, United States House of Representatives, [https://www.urban.org/sites/default/files/publication/101676/testimony\\_future\\_of\\_work\\_and\\_technology\\_-\\_jenny\\_yang\\_0\\_2.pdf](https://www.urban.org/sites/default/files/publication/101676/testimony_future_of_work_and_technology_-_jenny_yang_0_2.pdf).
13. New Zealand Government, Algorithm Assessment Report, October 2018, <https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf>.
14. New Zealand Law Foundation, Government use of Artificial Intelligence in New Zealand, 2019, [https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016\\_ILP\\_10\\_AILNZ-Report-released-27.5.2019.pdf](https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf).
15. AI Forum, Insights and information on the impacts of AI for New Zealand, <https://aiforum.org.nz/our-work/publications>.
16. Roger Brownsword, Professor of Law at King's College London, goes more deeply into the concept of social licence, discussing in a recently published paper the idea of having regulators include a triple licence for the regulation of any technology. This triple licence includes a global commons licence (regulators must respect the pre-conditions for human social existence), a community licence (regulators have a responsibility to respect the fundamental values of a particular community) and a social licence (seeking an acceptable balance of legitimate interests).
17. Taiuru, Karaitiana, Treaty of Waitangi/Te Tiriti and Māori Ethics Guidelines for: AI, Algorithms, Data and IOT, May 2020, <https://www.taiuru.maori.nz/wp-content/uploads/Treaty-of-Waitangi-Guidelines-for-a-Code-of-Ethics-using-Kaupapa-M%C4%81ori-Frameworks-for-AI-and-Algorithms-and-Data-Ktaiuru.pdf>.
18. "Higher Living Standards", The Treasury of New Zealand, <https://treasury.govt.nz/information-and-services/nz-economy/higher-living-standards>.
19. Pilot projects will also draw on global work on well-being, such as the IEEE's recent work on sustainability and well-being. For more info, visit: <https://standards.ieee.org/standard/7010-2020.html>.
20. Examples of these types of approaches include Denmark's Seal for IT use and data security. For more information, visit <https://eng.em.dk/news/2019/oktober/new-seal-for-it-security-and-responsible-data-use-is-in-its-way>, and New Zealand's Draft Algorithm Charter at <https://data.govt.nz/assets/Uploads/Draft-Algorithm-Charter-for-consultation.pdf>.
21. See, for example, Denmark: <https://en.digst.dk/policy-and-strategy/digital-ready-legislation/guidances-and-tools>.
22. "Reimagining Regulation for the Age of AI", World Economic Forum, <https://www.weforum.org/projects/reimagining-regulation-for-the-age-of-ai>.
23. "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems", IEEE, May 2019.
24. "Forty-two countries adopt new OECD Principles on Artificial Intelligence", OECD, May 2019, <https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.html>.
25. "Principles of Artificial Intelligence", Smart Dubai, <https://www.smartdubai.ae/initiatives/ai-principles>.

26. "On Artificial Intelligence - A European approach to excellence and trust", European Commission, 19 February 2020, [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf).
27. "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics", European Commission, 19 February 2020, [https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020\\_en\\_1.pdf](https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf).
28. Srivastava, Smriti, "The Three Pillars of European Approach to AI Excellence", Analytics Insight, 18 March 2020, <https://www.analyticsinsight.net/three-pillars-european-approach-ai-excellence>.
29. All Information (Except Text) for H.R.2231 - Algorithmic Accountability Act of 2019, <https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info#actionsOverview-content>.
30. Space limitations prevent us reviewing these developments in detail. See "U.S. AI Regulation Guide: Legislative Overview and Practical Considerations" by Baker McKenzie associate (and former WEF Fellow) Yoon Chae, in *Robotics, Artificial Intelligence & Law* / January–February 2020, Vol. 3, No. 1, pp. 17–40.
31. NITI Aayog, National Strategy for Artificial Intelligence, June 2018, [https://niti.gov.in/writereaddata/files/document\\_publication/NationalStrategy-for-AI-Discussion-Paper.pdf](https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf).
32. Office of the Privacy Commissioner of Canada, Consultation on the OPC's Proposals for ensuring appropriate regulation of artificial intelligence, 13 March 2020, [https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/consultation-ai/pos\\_ai\\_202001](https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/consultation-ai/pos_ai_202001).
33. The Conference toward AI Network Society, Draft AI R&D GUIDELINES for International Discussions, July 2017, [https://www.soumu.go.jp/main\\_content/000507517.pdf](https://www.soumu.go.jp/main_content/000507517.pdf).
34. Ministry of Industry, Business and Financial Affairs of Denmark, Denmark should be a frontrunner in responsible development and use of artificial intelligence (AI), March 2019, <https://eng.em.dk/news/2019/marts/denmark-should-be-a-frontrunner-in-responsible-development-and-use-of-artificial-intelligence-ai>.
35. Ministers for the Department of Industry, Science, Energy and Resources of Australia, Budget 2018 - New opportunities and jobs for Australian industry, May 2018, <https://www.minister.industry.gov.au/ministers/cash/media-releases/budget-2018-new-opportunities-and-jobs-australian-industry>.
36. Commonwealth Scientific and Industrial Research Organisation (CSIRO), Artificial Intelligence Australia's Ethics Framework - A Discussion Paper, 2019, [https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting\\_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf](https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf).
37. Calo, Ryan, "Artificial Intelligence policy: a primer and road map", SSRN, 19 October 2017, <https://ssrn.com/abstract+3015350>.
38. Selbst, Andrew, "Disparate Impact in Big Data Policing", SSRN, 17 April 2018, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2819182](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2819182).
39. Reed, Chris, "How should we regulate artificial intelligence?" The Royal Society, 6 August 2018, <https://doi.org/10.1098/rsta.2017.0360>.
40. For example, see their recent paper on developments in the US, <https://www.bakermckenzie.com/-/media/files/people/choe-yoon/rail-us-ai-regulation-guide.pdf>.
41. For example, see <https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusions>.
42. "Using personal information responsibly", Ministry of Social Development of New Zealand, <https://www.msd.govt.nz/about-msd-and-our-work/work-programmes/initiatives/phrae/index.html>.
43. Social Investment Agency, What you told us: Findings of the 'Your voice, your data, your say' engagement on social wellbeing and the protection and use of data, November 2018, <https://swa.govt.nz/assets/Uploads/what-you-told-us.pdf>.
44. Massey University, Shaping a framework for sharing personal data in NZ, <https://www.toiaria.org/ourdataourway>.
45. Human Rights and Technology, <https://tech.humanrights.gov.au>.
46. Clough, Ryan, "The Inevitability of AI Law and Policy: Preparing Government for the Era of Autonomous Machines", Public Knowledge, 2018, [https://www.publicknowledge.org/assets/uploads/blog/AI\\_Report.pdf](https://www.publicknowledge.org/assets/uploads/blog/AI_Report.pdf).
47. Clough, Ryan, "The Inevitability of AI Law and Policy: Preparing Government for the Era of Autonomous Machines", Public Knowledge, 2018, [https://www.publicknowledge.org/assets/uploads/blog/AI\\_Report.pdf](https://www.publicknowledge.org/assets/uploads/blog/AI_Report.pdf).
48. Wallach, Wendell, "Toward the Agile and Comprehensive International Governance of AI and Robotics", IEEE, 2019, <https://ieeexplore.ieee.org/document/8662741>.
49. Scherer, Matthew, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies", SSRN, 2016, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2609777](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2609777).
50. Government use of artificial intelligence in New Zealand: <https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf>
51. Erdelyi, Olivia Johanna and Judy Goldsmith, "Regulating Artificial Intelligence: Proposal for a Global Solution", 2 February 2018, AAAI/ACM Conference on AI, Ethics, and Society (AIES '18), February 2--3, 2018, New Orleans, LA, USA doi/10.1145/3278721.3278731. Available at SSRN: <https://ssrn.com/abstract=3263992>.
52. "9 recommendations to strengthen Danish businesses in the responsible use of data", Ministry of Industry, Business and Financial Affairs of Denmark, 22 November 2018, <https://eng.em.dk/news/2018/november/9-recommendations-to-strengthen-danish-businesses-in-the-responsible-use-of-data>.
53. Davenport, Thomas and Ravi Kalakota, "The potential for artificial intelligence in healthcare", *Future Healthcare Journal*, June 2019, 6(2), 94–98, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181>.

54. “Artificial Intelligence Promises a Personalized Education for All”, The Atlantic, 2017, <https://www.theatlantic.com/sponsored/vmware-2017/personalized-education/1667>.
55. Keats Citron, Danielle, “Technological Due Process”, Washington University Law Review, vol. 85, issue 6, 2008: [https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law\\_lawreview](https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview).
56. Mulligan, Deidre and Kenneth Bamberger, “Procurement As Policy: Administrative Process for Machine Learning”, Berkeley Technology Law Journal, vol. 34, 2019, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3464203](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3464203).
57. “Algorithmic Impact Assessment (AIA)”, Government of Canada, 2019, <https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html>.
58. World Economic Forum, UK Government First to Pilot AI Procurement Guidelines Co-Designed with World Economic Forum, [Press Release], 20 September 2019, <https://www.weforum.org/press/2019/09/uk-government-first-to-pilot-ai-procurement-guidelines-co-designed-with-world-economic-forum>.
59. Info-communications Media Development Authority (IMDA), Model Artificial Intelligence Governance Framework, second edition, 2020, <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.
60. “Privacy, Human Rights and Ethics framework”, Ministry of Social Development of New Zealand, <https://www.msd.govt.nz/documents/about-msd-and-our-work/work-programmes/initiatives/phrae/phrae-on-a-page.pdf>.
61. Raji, Inioluwa Deborah et al., “Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing”, Conference on Fairness, Accountability, and Transparency (FAT\* ’20), 27-30 January 2020, <https://arxiv.org/abs/2001.00973>.



---

COMMITTED TO  
IMPROVING THE STATE  
OF THE WORLD

---

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

---

**World Economic Forum**  
91–93 route de la Capite  
CH-1223 Cologny/Geneva  
Switzerland

Tel.: +41 (0) 22 869 1212  
Fax: +41 (0) 22 786 2744  
[contact@weforum.org](mailto:contact@weforum.org)  
[www.weforum.org](http://www.weforum.org)