



**CRIAT**

*Consortium for Robotics, Intelligence  
and Automation Technologies*

[www.criat.net](http://www.criat.net)

# **Transcending AI Policy Boundaries:**

Comparative Insights and the Case for Global Collaboration

**By:** Bruno Sousa  
Founder, The Consortium for Robotics, Intelligence and Automation Technologies

**Date:** June 29, 2023

# Executive Summary

We are at the dawn of a new age, and Artificial Intelligence (AI) is becoming ever more pervasive in our daily lives. Natural Language Processing, Language Models, and computer vision have enabled AI-driven systems to automate various tasks, from predictive auto-suggestions in emails to ChatGPT, a conversational AI. Moreover, AI-generated images and videos created through DALLE and MidJourney have opened up a world of unprecedented possibilities.

AI is also revolutionising how we work, with AI-powered chatbots replacing many customer service agents and AI-driven automation taking over mundane labour-intensive tasks. Moreover, various industries are applying AI-driven algorithms, from healthcare to finance, enabling them to find solutions to complex problems. As AI continues to shape our lives, from work to leisure, it is essential to recognise its potential and profound impact on our future.

Central to this exploration is the question: What measures and frameworks must be established and embraced globally for responsible, beneficial, and ethical governance of AI technologies? This pivotal inquiry guides our comparative analysis of the diverse AI governance strategies globally, highlighting the nuances.

In order to ensure the responsible and ethical use of Artificial Intelligence (AI), stakeholders must collaborate and cooperate. To this end, governments worldwide have invested heavily in AI research and development and have introduced regulations to govern the use of this technology. As a result, distinct approaches to AI governance have emerged from leading countries and regions, such as the United Kingdom, the United States, Canada, the European Union, China, Singapore, and India. Each nation and region has developed its own AI governance model that reflects its social, political, and economic context.

To guarantee the successful adoption of AI, we must take necessary steps to mitigate potential risks and financial impacts on employment while generating positive outcomes and respecting human rights. We as a society are ultimately responsible for assisting those affected by AI and considering labour regulations such as minimum wage or a minimum livable income.

When used correctly, AI can create more equitable and efficient systems, allocating resources fairly and providing economic opportunities for previously disadvantaged people. It is important to inform those in positions of authority and other key stakeholders about the potential risks associated with this opportunity and to guide them on managing them effectively.

This document aims to provide a strong base for achieving this by outlining risks and providing recommendations to encourage collaboration and cooperation between stakeholders.

# Table of Contents

- Executive Summary..... 2
- Table of Contents.....3
- Introduction..... 5
- AI Governance Frameworks..... 6
  - The European Union (EU): A Comprehensive Risk-Based Legal Framework.....6
  - The United Kingdom (UK): A Pragmatic and Pro-Innovation Framework..... 8
  - Canada: Equitable Coexistence Framework..... 9
  - The United States (US): A Multifaceted Framework..... 10
  - China’s: An Ambitious and State-Controlled Framework..... 11
  - Singapore: An Ethical and Commercially Oriented Framework..... 12
  - India: A Holistic and Socially Oriented Framework..... 13
- Comparison Summary..... 15
  - Policy Strategy..... 15
  - Ethical Guidelines..... 15
  - Government Involvement.....15
  - Public-Private Partnerships..... 16
  - AI Research and Development..... 16
  - Notable AI Policies..... 16
  - AI Laws/Regulations..... 16
- Analysis..... 17
- Risks and Challenges..... 19**
  - Risks..... 19
    - Technical Risks..... 19
    - Societal Risks..... 19
    - Long-term, Existential Risks..... 20
    - Risk Comparison Matrix..... 21
      - Technical Risks..... 21
      - Societal Risks..... 22
      - Long-term, Existential Risks..... 23
- Consensus and Alignment.....23
  - Diverse Ethical Standards and Cultural Differences..... 23
  - Disparate Regulatory Frameworks..... 24
  - Rapid Technological Advancement.....24
  - Economic Competition.....24
  - Transparency and Trust.....24
- Agreement on Important Definitions..... 24
  - Truth.....25

Self-Awareness and Consciousness.....	25
Sentience.....	25
Intelligence.....	25
Additional Concerns.....	26
Environmental Protection.....	26
Indigenous Rights.....	26
Autonomous Weapons Systems.....	26
<b>Recommendations.....</b>	<b>27</b>
Promote Transparency.....	27
Ethics in AI Development.....	27
Collaboration across Sectors.....	27
Invest in AI Literacy and Education.....	27
Develop Robust Security Measures:.....	28
Address the Long-term Risks of AI:.....	28
Unified International Guidelines:.....	28
Google’s AI Principles.....	28
The Asilomar AI Principles.....	29
OpenAI’s Charter.....	29
EU Ethics Guidelines for Trustworthy AI.....	29
IEEE’s Ethically Aligned Design.....	29
The Toronto Declaration.....	29
The Partnership on AI.....	29
Montreal Declaration for a Responsible Development of AI.....	30
OECD Principles on Artificial Intelligence.....	30
UNESCO’s recommendation on the ethics of AI.....	30
Establishing an International Agency for AI:.....	30
Inspiration from International Bodies.....	31
Fostering Consensus through Collaboration.....	31
Addressing Criticisms.....	32
<b>Conclusion.....</b>	<b>33</b>
<b>References.....</b>	<b>35</b>

# Introduction

We stand on the threshold of the “Fourth Industrial Revolution,” a term coined by Professor Klaus Schwab, founder and executive chairman of the World Economic Forum. This revolution, characterised by a fusion of technologies, blurs the lines between the physical, digital, and biological spheres. At its core is Artificial Intelligence (AI), a disruptive force with the potential to transform society in unprecedented ways.

The emergence of AI in 1956 has immensely impacted our lifestyles and working patterns. John McCarthy’s pioneering use of “Artificial Intelligence” or “AI” spurred tremendous research and development, resulting in significant advances in the field. After that, Frank Rosenblatt invented the Perceptron in the 1960s - the first neural network for computers - which paved the way for further progress. In the 1970s and 1980s, Stanford, MIT, and Carnegie Mellon achieved several breakthroughs. In 1997, IBM’s Deep Blue famously defeated world chess champion Garry Kasparov, and in the 2000s, we saw the development of self-driving cars by companies such as Tesla and Waymo. IBM Watson’s victory on Jeopardy! in 2011 was another remarkable feat, and more recently, DeepMind’s AlphaGo and OpenAI’s GPT-3 and GPT-4 language models have revolutionised natural language processing.

Artificial Intelligence (AI) technology is increasingly common in businesses, as it can improve productivity, enhance efficiency, and reduce costs. Despite the potential benefits, however, implementing AI has raised several ethical dilemmas and societal implications, which can harm fundamental rights, societal stability, and environmental protection. To ensure that the usage of AI is responsible and ethical, we must take necessary precautions, which include holding companies and individuals accountable for any misuse of AI and establishing legal frameworks by governments to protect citizens from potential abuses. Moreover, policymakers and businesses must consider the ethical implications of AI and ensure that they are using the technology appropriately.

We must work together to shape the future of Artificial Intelligence (AI) ethically and responsibly. In order to achieve this, it is essential to create a regulatory system that ensures AI is used safely, securely, and in an open manner. Moreover, we must create a governance system that benefits all stakeholders and encourages progress and innovation.

This document seeks to raise awareness and motivate readers to collaborate in creating a harmonious world of sustainability, technology, and responsible leadership. To that end, we must explore the strategies of major players such as the UK, EU, US, Canada, China, Singapore, and India to understand better how they manage AI. By doing so, we can construct an effective governance system that safeguards the interests of those involved.

In conclusion, we must recognise the power of collective action during this pivotal time to ensure a positive future for all. We must strive to establish a future of Artificial Intelligence (AI) that is fair, equitable, and provides a safe space for all to reach their potential. This framework should prioritise the safety and well-being of all citizens while guaranteeing that AI’s advantages are broadly shared. In order to accomplish this, we must collaborate, and each take

responsibility for our actions. We must develop and deploy efficient, ethical, and sustainable solutions. Moreover, we must be alert to AI's potential risks and dangers and ensure safeguards are in place to protect people, businesses, and society.

As a society, it is important to unite and make collective efforts towards creating a secure, beneficial, and responsible future for all. Let it be a legacy we can all be proud of, as it ensures a future of AI that is both prosperous and secure. Thus, this report invites all readers to join the noble pursuit of forging an inclusive and equitable future. Together, we can make a difference.

# AI Governance Frameworks

Integrating Artificial Intelligence (AI) into daily life has become inevitable in today's digitally accelerated world. However, the rapid adoption of AI technologies raises urgent questions regarding governance, ethical considerations, and societal impact. The exponential growth of AI demands a robust governance framework to ensure its ethical and safe usage.

Governance frameworks for AI are being developed and deployed globally, with key players being the European Union (EU), the United Kingdom (UK), Canada, the United States (US), China, Singapore, and India. Each framework brings its unique perspective, reflecting its socio-political environment and economic considerations. This section aims to provide a comparative analysis of these diverse AI governance frameworks, focusing on their strengths, weaknesses, and areas for improvement.

## The European Union (EU): A Comprehensive Risk-Based Legal Framework

The European Union has emerged as a global forerunner in creating a powerful legal framework for AI governance, leveraging several strategic initiatives over the years. Beginning with the EU's Declaration of Cooperation on Artificial Intelligence in 2018<sup>1</sup> and its Coordinated Plan on Artificial Intelligence in the same year<sup>2</sup>, the EU has consistently emphasised its commitment to comprehensive AI regulation.

In 2021, building on the foundational guidelines and plans, the European Commission proposed its first exhaustive legal framework on AI<sup>3</sup>. This framework outlines a risk-based approach, stratifying AI applications into four categories: "unacceptable risk," "high risk," "limited risk," and "minimal or no risk". Each category has its own set of obligations and requirements, reflecting the potential consequences and gravity of the application in question.

- **Unacceptable Risk:** These include AI applications deemed too harmful to be allowed, such as systems that manipulate human behaviour to circumvent users' free will and systems that permit 'social scoring' by governments<sup>3</sup>.
- **High Risk:** These include systems that must comply with strict conditions such as transparency, human oversight, risk assessment, dataset quality, result traceability, and robustness to ensure responsible usage<sup>3</sup>. Applications in this category include critical infrastructures, educational or vocational training systems, safety components of products, and employment and workers management<sup>3</sup>.

- **Limited Risk:** AI applications in this category, such as chatbots, are subject to specific transparency obligations. Users must know they interact with a machine to make an informed decision <sup>3</sup>.
- **Minimal or No Risk:** Most AI systems fall into this category, and therefore, applications like AI-enabled video games or spam filters can operate freely without additional regulatory intervention <sup>3</sup>.

The EU's framework places substantial emphasis on regulating high-risk AI systems. This regulatory stance aligns with the EU's Ethical Guidelines for Trustworthy AI, published in 2019 by the High-Level Expert Group on AI. It stressed transparency, human agency, and oversight among its key requirements <sup>4</sup>. The EU's focus on resilience against errors in AI systems, especially those that continue learning post-deployment, highlights its commitment to AI safety and reliability.

This risk-based approach to AI governance goes beyond technical considerations. It outlines clear responsibilities for providers and users of high-risk AI systems, including immediately reporting serious incidents or breaches of fundamental rights <sup>3</sup>. To ensure the consistent application of these regulations, the EU also proposed establishing a 'European Artificial Intelligence Board' <sup>3</sup>.

In its broader strategy for data, the EU has proposed creating single market data spaces, including one for AI <sup>5</sup>, demonstrating its proactive approach to governance. This thorough, risk-based approach reflects a conscious effort to address AI-specific risks preemptively, position Europe as a global leader, and maintain transparency, human oversight and a high safety standard.

The European Union's comprehensive legal framework demonstrates a lesson learned from past experiences, such as data privacy, where the lack of early, all-encompassing regulation led to several high-profile violations and controversies. By establishing a robust AI governance framework early on, the EU aims to avoid similar pitfalls, foster trust in AI systems, and ensure the alignment of AI development with societal values and legal principles.

The EU's framework is an instructive model for globally harmonised AI governance. Margrethe Vestager, Executive Vice President of the European Commission for A Europe Fit for the Digital Age, underlines the critical role of trust in adopting and implementing AI technologies. In a powerful statement, she emphasises, "On Artificial Intelligence, trust is a must, not a nice to have... By setting the standards, we can pave the way to ethical technology."

This thinking exemplifies the European Union's dedication to utilising AI and automation in a transparent, accountable, and human rights-oriented manner. Vestager's words encapsulate the EU's regulatory approach towards AI, which strongly emphasises establishing ethical and robust standards to guide AI development and use. This regulatory foresight positions the EU as a global leader in shaping the ethical trajectory of AI, demonstrating the belief that societal trust



is crucial for the acceptance of AI and integral to its successful integration into various aspects of daily life.

## The United Kingdom (UK): A Pragmatic and Pro-Innovation Framework

With the backdrop of Brexit, the United Kingdom (UK) charted its unique course in AI governance, taking a markedly pro-innovation stance <sup>7</sup>. The UK diverged from its European counterparts, eschewing the development of an AI-specific regulatory framework. Instead, as articulated in its 2022 AI White Paper, the UK adopted a strategy delegating the handling of AI-specific risks to pre-existing regulatory bodies <sup>7</sup>.

This governance model, driven by innovation, is designed to encourage technological progression and economic growth while safeguarding core societal values like safety, privacy, and human rights <sup>7</sup>. Its flexibility characterises the UK's model, emphasising AI's application more than the technology itself <sup>7</sup>. This adaptability allows for swift responses to the rapid evolution of AI technologies and applications <sup>7</sup>.

However, the UK's approach does not ignore potential AI risks and challenges. The White Paper acknowledges the need for thorough cross-sector risk assessments, especially concerning "high impact but low probability" existential threats that could be posed by Artificial General Intelligence (AGI) <sup>7</sup>. Such threats, though unlikely, could destabilise democratic institutions through disinformation or foster systems that violate existing laws such as the Equality Act 2010 <sup>8</sup> or the Human Rights Act 1998 <sup>9</sup>.

Despite these proactive measures, critics express concern over the UK's reliance on industry-led, voluntary AI standards <sup>7</sup>. The apprehension is that this strategy may fail to sufficiently safeguard consumer rights and societal interests, possibly prioritising economic growth over ethical considerations <sup>7</sup>.

Reiterating the UK's commitment to fostering an AI-friendly regulatory environment, the UK AI Council unveiled the AI Roadmap in 2021 <sup>10</sup>. This roadmap emphasises the balance between technological progression and ethical governance, underlining the importance of vigilance over AI's ethical implications <sup>10</sup>.

Sir Nigel Shadbolt, a leading figure in AI in the UK, encapsulates the essence of the UK's AI governance in his quote: "We need digital infrastructures that will promote a plurality of methods of data sovereignty and governance instead of imposing a single policy fits-all platform governance model, which has strained and undermined the ability for governments to protect and support their citizens' digital rights." <sup>11</sup>.

# Canada: Equitable Coexistence Framework

Canada is among the pioneers leading the global AI revolution. It has a robust and dynamic AI ecosystem featuring prominent research centres, startups, and collaborative projects with technology giants. While the nation aspires to become an AI hub, it also grapples with the formidable task of AI governance, seeking to balance economic gains with societal norms and values<sup>12</sup>.

Canada is developing its AI governance model, and legislative efforts are underway but have not been formally enacted. These efforts are pivotal to ensuring AI's impact on Canadian society is beneficial and risk mitigated. The proposed bills, in various stages of the legislative process, are designed to address the complexity of AI regulation, encompassing consumer privacy, data protection, fairness, accountability, and overall societal impact.

Bill C-427, for example, sought to make agriculture and agri-food sectors innovation hubs, with a specific focus on AI and robotics<sup>13</sup>. However, it currently remains outside the Order of Precedence. Similarly, the ambitious Bill C-27, the Digital Charter Implementation Act, is a consideration in committee in the House of Commons<sup>14</sup>. This act aims to regulate AI systems' trade, requiring certain individuals to adopt measures to mitigate harm risks and biased outputs from high-impact AI systems. Nevertheless, critics have heavily condemned this act for its narrow scope, which only applies to the private sector.

In its 2019 Directive on Automated Decision-Making, the Canadian government outlined a proposed vision for AI usage centred on the principles of transparency, accountability, legality, and procedural fairness<sup>15</sup>. It also incorporated Algorithmic Impact Assessments (AIAs) for all government-automated decision-making applications.

Canada is considering Bill C-26, or the Act Respecting Cyber Security (ARCS)<sup>16</sup>, in its ongoing legislative agenda. This legislation would focus on stringent cybersecurity measures to safeguard vital systems and services, reinforcing the government's commitment to protecting the nation's cyberspace and critical infrastructure.

Finally, Bill C-43, intended to support a pan-Canadian AI strategy through substantial funding, has yet to pass its second reading in the House of Commons<sup>17</sup>. The bill proposed a significant investment in the Canadian Institute for Advanced Research to bolster Canada's national AI strategy.

As Yoshua Bengio, a renowned Canadian computer scientist and a global figure in AI research, believes that the Canadian Government is moving "way too slowly"<sup>18</sup> and recently stated, "I agree with the pause. However, I don't think it should be just OpenAI, and I don't think it should be only the United States, so it's going to take more time, but I really think that it has to move to the international arena as quickly as possible."<sup>19</sup>

In line with this vision, Canada's AI governance framework continues to evolve, acknowledging the necessity to balance the economic ambitions associated with AI with the equally important societal values at stake.

## The United States (US): A Multifaceted Framework

The landscape of artificial intelligence (AI) governance in the United States is a multi-dimensional tapestry of evolving policy, regulation, and practical applications. Despite its complex and, at times, inconsistent approach, the US is unequivocally committed to leveraging the potential of AI while addressing its societal implications.

The US Congress has displayed significant strides in the field, passing 36 bills concerning AI between 2021 and 2022, of which 11 became law. These bills range from the AI Training Act (S.2551) [20](#), which aims to enhance AI competence among federal employees, to the Infrastructure Investment and Jobs Act (H.R.3684) [21](#), focusing on AI's role in technology and infrastructure development.

The varied nature of these bills demonstrates an understanding of AI's broad impact and potential. State-level efforts are no less important. California's Automated Decision Systems Accountability Act of 2020 [22](#), for instance, underscores the state's commitment to transparency and fairness in AI. Nevertheless, the lack of a unified national approach to AI governance challenges harmonising regulations across different states and sectors.

Recent Senate hearings in May 2023 highlighted the growing consensus around stronger AI control mechanisms. OpenAI CEO Sam Altman, whose company created the ChatGPT and DALL-E text and image generation tools, thinks that "regulatory intervention by governments will be critical to mitigate the risks" [23](#). In his opening remarks at a recent Senate Judiciary subcommittee hearing, he made it clear that it has to be an all-stakeholder, universal scenario when he stated, "...we believe we can and must work together to identify and manage the potential downsides so that we can all enjoy the tremendous upsides. It is essential that powerful AI is developed with democratic values in mind..." [23](#).

Efforts to create a framework for AI governance are also emerging at the federal level. The White House Office of Science and Technology Policy (OSTP) is drafting a "Bill of Rights for an AI-powered world" [24](#), an initiative that symbolises the US' commitment to mitigating AI's potential threats while preserving democratic principles.

At the same time, the National Institute of Standards and Technology (NIST) is developing an AI risk management framework [25](#). The framework expects to establish standardised practices and language for the development of AI technology, with a focus on critical attributes like accuracy, security, and accountability.

Notable professionals in the field echo the progress in the US AI governance landscape. Dr Fei-Fei Li, co-Director of Stanford's Human-Centred AI Institute, has emphasised, "AI is a story about humanity, the human values we want our technology to embody, the people who develop it, and the communities it affects." <sup>26</sup>.

In conclusion, while the American AI governance landscape is rich and evolving, there is a pressing need for a unified national strategy that balances innovation, regulation, and the protection of civil liberties. By harmonising these elements, we can ensure AI serves the collective good of humanity, regardless of geographical boundaries.

## China's: An Ambitious and State-Controlled Framework

China's AI governance narrative is marked by ambition and stringent state control, as articulated in its "New Generation Artificial Intelligence Development Plan" <sup>27</sup>. The nation seeks to solidify its status as a global AI leader by 2030 <sup>27</sup>. As this development plan outlines, the Chinese strategy highlights a distinctive blend of state intervention emphasising the security, controllability, and robustness of AI systems <sup>27</sup>.

Recognising AI as a strategic asset in pursuing global technological dominance, the Chinese government's approach to AI governance is state-centric, marked by overarching government policies and programs that guide AI development <sup>28</sup>. This hands-on approach, however, has not been without its share of controversy, with concerns over state surveillance and control gaining traction on the international stage <sup>29</sup>.

Despite these concerns, China's focus on AI system robustness and security is noteworthy. The government's measures to ensure the controllability of AI technologies underline the crucial role national security plays in AI governance <sup>27</sup>. Furthermore, recognising AI's transformative power signifies understanding the risks and challenges posed by AI in an increasingly digital landscape <sup>30</sup>.

However, China's approach extends beyond mere security. As articulated in the Beijing AI Principles, there is a prominent focus on promoting harmony, friendliness, and fairness in AI systems <sup>31</sup>. These principles underscore the nation's emphasis on social stability, demonstrating the recognition of AI as a potent tool capable of driving societal change. On the practical side, Chinese technology companies, such as Baidu, Alibaba, and Tencent, are trailblazers in AI research and development <sup>30</sup>. Their efforts, from e-commerce and social media applications to autonomous driving and healthcare, bolster China's ambitious AI objectives.

Kai-Fu Lee, CEO of Sinovation Ventures and former President of Google China, echoes China's aspiration to become the world's leading AI superpower. In his book, he writes, "Meeting these challenges means we cannot afford to passively react. We must proactively seize the opportunity that the material wealth of AI will grant us and use it to reconstruct our economies and rewrite our social contracts." <sup>32</sup> underscoring that financial gain is China's primary objective to leverage AI. He also emphasised, "Behind these efforts lies a core difference in American and Chinese political culture: while America's combative political system aggressively punishes missteps or waste in funding technological upgrades, China's techno-utilitarian approach rewards proactive investment and adoption." <sup>32</sup> suggesting that the very structure of governance in China will allow for blazing-fast AI integration, which may lead them to become an AI supremacy.

In summary, while China's AI governance approach invites international concerns, its strategic focus on AI as an instrument for national development and social stability is significant. As evidenced by the country's AI development plan, the balance between state control and ethical considerations in AI use will continue to shape China's AI governance landscape <sup>27</sup>.

## Singapore: An Ethical and Commercially Oriented Framework

Singapore, a city-state renowned for its technological prowess and innovative spirit, has made substantial strides in crafting a robust and pragmatic AI governance framework. Singapore's approach to AI governance melds ethical concerns with commercial innovation, creating a dynamic environment that enables businesses to leverage AI while respecting ethical considerations and data privacy regulations.

In 2019, the Infocomm Media Development Authority (IMDA) of Singapore published the second edition of its Model AI Governance Framework <sup>33</sup>, an in-depth guidance document designed to assist organisations in implementing AI solutions responsibly. Although not legally binding, this framework indicates the government's perspective on ethical AI use, focusing on principles such as explainability, transparency, and fairness in AI decision-making. The model also promotes active human involvement in AI-augmented decisions, reinforcing the centrality of human oversight in AI applications <sup>33</sup>.

Notably, this framework is one of the first in the world to provide comprehensive, implementable guidance on AI, incorporating case studies and checklists to facilitate risk identification and mitigation. The IMDA has also set up an Advisory Council on the Ethical Use of AI and Data, tasked with assisting the government in developing ethical standards and guidelines for AI applications <sup>34</sup>.

The Personal Data Protection Act (PDPA) indirectly impacts the deployment of AI in Singapore <sup>35</sup>. The PDPA enforces responsible collection, use, and disclosure of personal data, imposing obligations on AI systems that process such information. To further assist organisations in managing AI and data analytics concerning the PDPA, the IMDA and the Personal Data Protection Commission (PDPC) have jointly released a guide <sup>35</sup>.

On the commercial front, Singapore's proactive AI stance is evident in the vibrant AI ecosystem, which includes local startups, global tech giants, and research institutions. For instance, the city-state is home to Sea Group, Grab, and Lazada, which extensively use AI in their operations <sup>36</sup>.

In the concluding statements of a keynote address by Mr S Iswaran, Minister for Communications and Information in Singapore, during the Asia House Global Trade Dialogue in November of 2019, Iswaran asserts both the potential benefit and risks of AI, saying, "technological advances and the digital economy present significant new opportunities. They also pose profound disruptive threats to businesses, jobs and industries." <sup>37</sup>. Followed by a realisation that "To mitigate the risks of dislocation and ensure an equitable distribution of outcomes and the access to them, we must be inclusive in our approach - that means governments taking a detailed and comprehensive approach and collaborating with unions and industry" <sup>37</sup> alluding to a greater sense of organisational effort in AI governance.

In conclusion, Singapore's pragmatic and balanced approach to AI governance creates an environment conducive to commercial innovation and ethical AI use. It offers a valuable model for other nations looking to navigate the complexities of AI governance.

## India: A Holistic and Socially Oriented Framework

India's AI journey involves strategic planning, fostering technological innovation, and adopting a socially oriented approach. India's AI policy-making has been structured through the National Strategy for Artificial Intelligence (NSAI) <sup>38</sup>, emphasising AI utilisation in various sectors. The government recognises the transformative power of AI, envisioning it as a tool for social good and an enabler for inclusive economic growth.

The government has dedicated significant efforts to cultivating a strong AI ecosystem within the country. Collaboration with industry players, startups, and academic institutions is a cornerstone of this effort <sup>39</sup>. By nurturing these partnerships, India aims to foster AI innovation across both the private and public sectors, ensuring widespread, impactful deployment of AI technologies.

Funding for AI research and development is integral to India's AI strategy. Establishing several AI Centers of Excellence nationwide underlines the government's commitment in this regard <sup>40</sup>. These centres are hubs for advanced AI research and the development of AI solutions tailored to India's unique socio-economic conditions.

In its quest to harness the power of AI for societal upliftment, India has implemented various AI-based initiatives to enhance public welfare. For example, the government has deployed AI technologies to boost agricultural productivity, improve healthcare outcomes, and enrich educational resources <sup>41</sup>. Such efforts highlight India's commitment to employing AI for societal transformation.

To capitalise on the vast potential of AI, India also recognises the importance of nurturing a skilled workforce. The government has launched several skill development and AI education programs to equip the nation's youth with the necessary AI skills <sup>42</sup>. Such initiatives foster a talent pool capable of steering India's AI future and establishing the country as a global AI powerhouse.

International collaboration is another key element of India's AI strategy. By seeking to learn from global best practices and encouraging cross-border innovation, India is committed to leveraging AI as a tool for sustainable and inclusive growth <sup>43</sup>.

In conclusion, India's emerging AI governance framework is marked by comprehensive planning, a focus on research and development, and a strong emphasis on using AI as a catalyst for social good. The government's ongoing efforts suggest a promising path toward harnessing the full potential of AI for national development and societal prosperity.

# Comparison Summary

## Policy Strategy

US	UK	EU	Canada	India	China	Singapore
Fragmented, market-led approach	Coordinated approach with AI Sector Deal	Comprehensive Strategy with AI White Paper	The balance between commercial application and ethical considerations	National Strategy for AI emphasising social good	Ambitious, state-centric approach	The balance between commercial application and ethical considerations

## Ethical Guidelines

US	UK	EU	Canada	India	China	Singapore
Private sector-led, with varying adherence	AI Council guidelines; public-private cooperation	Emphasis on trustworthiness and human rights	Focus on human-centric AI	Societal transformation emphasis	Beijing AI Principles focus on harmony, friendliness, fairness	Model AI Governance Framework emphasising fairness, ethics, human-centric values

## Government Involvement

US	UK	EU	Canada	India	China	Singapore
Limited; main efforts driven by private sector	Significant, with AI Council and Office for AI	Strong, with coordinated efforts at EU level	Significant government involvement in policy making and research funding	Central role of government in policy making and research funding	High level of state control and intervention	Strong, with coordinated efforts at government level



## Public-Private Partnerships

US	UK	EU	Canada	India	China	Singapore
Extensive, especially with tech giants	Extensive, via AI Sector Deal	Encouraged under Coordinated Plan	Extensive, especially through CIFAR, AI hubs	Encouraged under National AI Strategy	Major tech firms deeply integrated in national strategy	Strong collaborations, particularly with Smart Nation initiative

## AI Research and Development

US	UK	EU	Canada	India	China	Singapore
World-leading with tech giants and top universities	Significant, with Alan Turing Institute playing a central role	Significant funding for research and innovation	Strong, especially with CIFAR, VECTOR Institute	Emphasis on AI centres of excellence	Major tech firms at forefront of AI R&D	Emphasis on research institutes and AI technology adoption

## Notable AI Policies

US	UK	EU	Canada	India	China	Singapore
Executive Order on Maintaining American Leadership in AI <sup>44</sup>	AI Sector Deal <sup>45</sup>	AI Coordinated Plan <sup>47</sup>	Pan-Canadian AI Strategy <sup>49</sup>	National Strategy for Artificial Intelligence <sup>38</sup>	New Generation Artificial Intelligence Development Plan <sup>27</sup>	National AI Strategy <sup>51</sup>

## AI Laws/Regulations

US	UK	EU	Canada	India	China	Singapore
None at federal level, some <sup>22</sup> at state level	Data Protection Act <sup>46</sup> impacts AI deployment	AI Act <sup>3</sup>	Personal Information Protection and Electronic Documents Act <sup>50</sup> impacts AI	None at national level, some initiatives at state level	Cyber-security Law <sup>52</sup> impacts AI deployment	Personal Data Protection Act <sup>53</sup> impacts AI deployment

# Analysis

With this comprehensive exploration of AI governance globally, it is evident that AI, as an international phenomenon, is governed by a patchwork of unique regional approaches and strategies. These approaches reflect the complex interplay between AI technology and the societal, economic, legal, and ethical dimensions within each region.

The United States operates under a market-led governance strategy, where tech giants and leading universities contribute significantly to AI policy and practices, indicating the profound influence of the private sector on AI's trajectory. The vibrant dynamism of the American AI industry is palpable, even without a unifying national policy.

The UK and the European Union have vigorously championed ethical considerations in AI governance. The EU's proposed AI Act is a significant leap towards an ethical AI future. Both regions highlight the role of the government in directing AI development and have made significant investments in AI research, innovation, and commercialisation.

Canada's approach to AI governance mirrors its European counterparts, balancing commercial application with ethical considerations and adopting a people-centric approach. Major corporations in the tech sector contribute significantly to AI research and applications, shaping the national AI landscape while adhering to government-led ethical guidelines.

India's AI governance emphasises societal transformation, where government initiatives and policy-making are complemented by robust participation from the private sector in AI-driven solutions targeting diverse societal needs.

China's approach to AI governance reflects a state-centric model, with AI considered a strategic asset interweaving national security, economic growth, and innovation. Nevertheless, China's tech giants also have a significant role in advancing AI technology, albeit within the framework of stringent state control.

Singapore, known for its Smart Nation initiative, represents a small city-state's ability to leverage AI effectively, balancing commercial and ethical considerations. Like the EU and Canada, Singapore's private sector is deeply involved in shaping the nation's AI landscape while adhering to a strong ethical code.

While not exhaustive, this comparison provides a snapshot of the diverse approaches taken by leading global players in AI. Each country or region has made significant strides in AI, contributing valuable insights into various AI strategies, policies, and frameworks globally.

Progress in AI governance extends well beyond these regions. Countries like Japan, South Korea, Australia, and nations in the Middle East and Africa are also making significant advancements in AI. They are carving out their unique positions in the global AI ecosystem, often led by ambitious private sector companies that align their AI strategies with national objectives.

Additionally, global political parties, like the Global Greens, have begun to acknowledge the significance of AI governance. In June 2023, during the Global Greens Congress 2023 in Korea, an Emergency Resolution for the Effective Regulation of AI Technology was endorsed [119](#). The resolution, proposed by the Green Party Korea and Green Party Indonesia, emphasises the responsible development and use of AI, mandatory legal compliance, and the establishment of frameworks promoting safety, transparency, and accountability in AI. Moreover, it calls for minimising AI's environmental impact and ensuring justice for workers in AI development. This crucial step from a major global political consortium reiterates the need for international collaboration in shaping the future of AI, aligning it with principles of democracy, sustainability, and social good.

In conclusion, each region has recognised AI's transformative power and its societal, economic, and ethical implications. These diverse frameworks underscore AI's complex and multifaceted nature, necessitating continued dialogue and idea-sharing across borders. Despite unique strengths and features, all governance frameworks can improve. Commercial entities in each country are shaping AI policy frameworks while trying to balance innovation and ethical considerations. The path to global alignment requires consensus-building and adaptability to accommodate rapidly evolving AI technologies, societal needs, and ethical considerations.

# Risks and Challenges

## Risks

Artificial Intelligence, heralded as the driving force behind the next industrial revolution, carries potential benefits and a range of substantial risks. These risks span technical and societal domains, from security vulnerabilities and black-box decision-making processes to job displacement and privacy concerns. Mitigating these risks necessitates keeping safety and ethics at the forefront of AI development, involving robust security measures, ethical guidelines and regulations, and emphasising transparency and accountability.

### Technical Risks

One category of AI risks falls within the technical domain. These include security vulnerabilities, reliability and robustness issues, and lack of transparency, often called the black-box problem.

- 1. Security Vulnerabilities:**

AI systems, like any digital technology, are susceptible to cyberattacks. These can include data manipulation, adversarial attacks, and system hacking, compromising the integrity of AI systems and potentially leading to harmful outcomes [54](#).

- 2. Reliability and Robustness:**

AI systems might perform excellently in their training environment but fail to generalise when encountering different situations [55](#). This unpredictability can be particularly concerning in healthcare or autonomous vehicles, where system errors can result in detrimental outcomes.

- 3. Lack of Transparency:**

Deep learning systems often suffer from a lack of interpretability, which can cause difficulties in understanding why an AI system made a specific decision [56](#). This opacity can lead to trust and acceptance issues among users.

### Societal Risks

AI also presents risks with broader societal implications, including job displacement, privacy concerns, biases and discrimination, economic inequality, and misuse of AI.

- 1. Job Displacement:**

The potential for AI to automate jobs has stirred debate, sparking concerns over significant job displacement across various industries [57](#).

## 2. Privacy Concerns:

AI systems, particularly those involved in data analysis and predictive analytics, can infringe on privacy rights, as they often require processing large amounts of personal data [58](#).

## 3. Bias and Discrimination:

AI systems can inadvertently perpetuate or amplify biases in their training data or design, leading to discriminatory outcomes [59](#).

## 4. Economic Inequality:

The benefits of AI could be unevenly distributed, potentially exacerbating economic inequalities [60](#).

## 5. Misuse of AI:

AI can be employed in harmful ways, whether malicious actors create “deep fakes” or develop lethal weapons [61](#).

## Long-term, Existential Risks

Some researchers have voiced speculative concerns over what is considered the long-term existential risks, such as the potential for AI to exceed human Intelligence in all areas, leading to the creation of a super-intelligent entity [62](#). This entity’s motivations and goals could be misaligned with human interests in such a scenario.

AI or AGI’s malicious misuse to extract and use sensitive information presents a significant concern, ranging from data breaches and identity theft to automated hacking and AI weaponisation [63](#), [64](#), [65](#), [66](#), [67](#). Similarly, the application of AI in enhancing traditional weaponry and military technology raises serious ethical and security concerns [68](#), [69](#), [70](#), [71](#), [72](#), [73](#).

AI and machine learning’s potential application in the design and synthesis of harmful substances represents another significant risk. AI could be used in developing chemical or biological weapons, synthesising harmful or illegal substances, and environmental hazards [74](#), [75](#), [76](#), [77](#).

While these risks are substantial and require careful attention, it’s essential to remember that they are primarily hypothetical at this point, often associated with the notion of superintelligence. They depend on a combination of AI capability and malicious intent. Furthermore, such uses would likely contravene multiple laws and international treaties.

However, it is crucial not to allow these speculative scenarios to overshadow more immediate and tangible concerns related to AI development and use. Recognised researchers in AI ethics have highlighted crucial issues such as bias, lack of diversity and inclusion, and the need for transparency in AI research [120](#). These pressing issues demand our attention right now to ensure that AI technologies are developed and used responsibly, fairly, and to benefit all of society.

## Risk Comparison Matrix

As we embark on an exploration of artificial Intelligence (AI) and its associated risks, it is pertinent that we leverage an established framework to guide our understanding. For this purpose, we employ the risk level classifications proposed by the European Commission <sup>3</sup>. This framework creates an insightful backdrop against which AI’s potential risks can be assessed. The following risk matrix will utilise this risk level system to present a structured and comprehensive assessment of AI’s potential risks.

### Technical Risks

Risk Level/ Category	Low Risk	Medium Risk	High Risk	Unacceptable Risk
<b>Data Privacy and Security</b> <sup>81</sup>	Unauthorised access to non-sensitive data	Data breaches involving personal information	Large-scale data breaches, leading to identity theft	Widespread and systemic data breaches undermining national security
<b>Algorithmic Bias</b> <sup>81</sup>	Minor bias in non-critical systems	Bias in commercial recommendation algorithms	Discrimination in high-impact areas like job screening	Systemic bias leads to widespread discrimination and social harm
<b>Autonomous Systems</b> <sup>78</sup>	Minor accidents caused by autonomous systems	Major accidents caused by autonomous vehicles	Large-scale accidents causing widespread damage and loss of life	Autonomous weapons causing uncontrolled destruction

## Societal Risks

<b>Risk Level/ Category</b>	<b>Low Risk</b>	<b>Medium Risk</b>	<b>High Risk</b>	<b>Unacceptable Risk</b>
<b>Job Displacement</b> <a href="#">80, 82</a>	AI replacing tasks within jobs, leading to job evolution	AI replacing certain jobs, leading to retraining	AI causing widespread job loss in certain industries	AI is causing massive, widespread job loss across multiple industries
<b>Malicious Use of AI</b> <a href="#">79</a>	AI used to create minor harmful substances	AI used to synthesise illegal drugs	AI used in the production of harmful biological or chemical agents	AI used in the large-scale creation and release of environmentally harmful substances
<b>Impact on social interactions</b> <a href="#">81, 82</a>	Minimal AI applications on social platforms	Increased use of AI in social media platforms leads to echo chambers and misinformation	Extensive use of AI led to significant changes in human behaviour and societal norms	AI applications resulting in widespread social disruption or societal harm

## Long-term, Existential Risks

Risk Level/ Category	Low Risk	Medium Risk	High Risk	Unacceptable Risk
<b>Super Intelligence and loss of control</b> <sup>62, 78, 82</sup>	Limited AI applications that augment human capabilities	The development of narrow AI systems that can outperform humans in specific tasks	Development of Artificial General Intelligence (AGI) without appropriate safeguards	Uncontrolled superintelligence, surpassing human abilities with potential existential risks
<b>Creation of toxic substances</b> <sup>79</sup>	AI used for beneficial chemical or biological synthesis	AI assists in creating potentially harmful substances without proper regulations	AI used to design and produce harmful substances in large quantities	AI misuse leads to large-scale production of highly toxic substances
<b>Weaponization of AI</b> <sup>71, 82</sup>	Use of AI for petty crime	AI used for cybercrime, such as ransomware	AI used for major cyber attacks on infrastructure	Use of AI in the creation of weapons of mass destruction
<b>Energy consumption and environmental impact</b> <sup>83</sup>	AI applications with minimal energy consumption	Moderate energy use in data centres and other AI infrastructure	Significant energy use leading to environmental harm	AI operations leading to severe environmental damage and unsustainable energy consumption

## Consensus and Alignment

The AI industry is marked by several consensus and alignment issues that stem from a multitude of factors, including varying ethical standards, different regional regulations, the rapid pace of technological advancement, and economic competition.

### Diverse Ethical Standards and Cultural Differences

A significant source of consensus and alignment issues in the AI industry revolves around varying ethical standards and cultural differences. What one culture or society deems ethical or acceptable regarding AI use may be seen differently in another. This divergence extends to privacy norms, decision-making bias, and the balance between security and personal freedoms <sup>84</sup>.



## Disparate Regulatory Frameworks

AI is a global phenomenon, but regulations are regional or national. Countries worldwide have different laws and regulations regarding AI and its applications. While some countries have extensive AI regulations, others have limited or no specific laws addressing AI. This disparity in regulatory frameworks can lead to alignment issues, especially for multinational corporations operating across multiple jurisdictions <sup>85</sup>.

## Rapid Technological Advancement

The rapid pace of AI technology development often outstrips the ability of societies and regulators to understand and govern its implications fully. As such, there can be a lag in policy and regulatory responses to new AI advancements, leading to a need for more consensus on how to address these emerging technologies <sup>86</sup>.

## Economic Competition

Economic competition is another significant factor contributing to alignment issues. Companies and countries alike are in a race to gain a competitive edge in AI, which can lead to discrepancies in standards and practices. Without universally agreed-upon regulations, individual entities may prioritise short-term economic gain over long-term ethical and societal considerations <sup>87</sup>.

## Transparency and Trust

Lastly, transparency and trust contribute to the need for more consensus in the AI industry. As AI systems become increasingly complex, it can take time to understand how they make decisions. This lack of transparency can hinder efforts to align standards and practices, as stakeholders may have different levels of trust in these systems based on their understanding and acceptance of AI <sup>88</sup>.

## Agreement on Important Definitions

Less tangible but equally important concepts like “truth,” “self-awareness,” “consciousness,” “sentience,” and “intelligence” are central to AI discussions. However, they are also areas of considerable debate and lack of agreement. These topics are deeply philosophical and psychological, with different interpretations across cultures, philosophical schools of thought, and scientific disciplines.

## Truth

In AI, truth pertains to the reliability and accuracy of an AI system's data, conclusions, or actions. However, defining truth is complex, even in human terms, as it often relies on perception, context, and subjectivity. The alignment issue arises when we consider that an AI system's understanding of 'truth' is determined by the data it is trained on, which can contain biases or inaccuracies [89](#).

## Self-Awareness and Consciousness

There is significant debate about whether AI can truly achieve self-awareness or consciousness. These terms denote a level of introspection and subjective experience that, as far as we know, require biological processes not present in AI. Even among humans, the nature of consciousness and self-awareness is a contentious philosophical and scientific issue, with various theories but no universally agreed-upon definitions [90](#).

## Sentience

Sentience refers to the capacity to have subjective experiences and feelings or the ability to perceive and respond to sensations. In animals, sentience is generally associated with conscious awareness and the capacity to experience pleasure and pain. However, this concept becomes contentious in AI as current AI needs the biological structures (like a nervous system) that facilitate sentience in living organisms [91](#).

## Intelligence

Intelligence, too, is a challenging concept to define, even within human contexts. Generally, it refers to the ability to learn, understand, and apply knowledge, reason abstractly, adapt to new situations, comprehend complex ideas, etc. In the realm of AI, Intelligence typically refers to a system's ability to perform tasks that would usually require human Intelligence—this includes tasks like understanding natural language, recognising patterns, solving problems, and making decisions. However, while we may describe an AI as intelligent, there is still disagreement on whether this equates to human-like Intelligence, as AI lacks traits like consciousness and emotional understanding [92](#).

In summary, the consensus and alignment issues in the AI industry are multifaceted, rooted in diverse ethical standards, varying regional regulations, rapid technological advancements, economic competition, and issues around transparency and trust. These complexities underline why reaching a consensus on these topics is challenging. Addressing these challenges will require concerted, collaborative efforts from governments, industry, academia, and civil society worldwide.

## Additional Concerns

As we delve deeper into Artificial Intelligence (AI), it becomes increasingly apparent that its influence extends beyond technology and innovation. AI is not merely a tool but a transformative phenomenon with far-reaching implications for our environment, indigenous rights, and even warfare. This section delves into these areas, highlighting AI's potential pitfalls and challenges and suggesting possible strategies to mitigate them.

### Environmental Protection

AI has a significant environmental footprint, with large-scale AI models requiring extensive computational power and energy. This requirement leads to substantial carbon emissions, thereby contributing to climate change. Current EU and UK AI governance frameworks need to address this issue directly instead of leaning on existing regulations and voluntary codes of conduct <sup>93</sup>. Addressing the environmental impact of AI is not merely a necessity for our planet's health but also a potential avenue for AI's positive impact. By incorporating environmental sustainability into AI systems' design, operation, and decommissioning, we can mitigate their environmental footprint while leveraging AI to promote sustainability <sup>94</sup>.

### Indigenous Rights

The respect for indigenous rights and territories is another crucial area where AI intersects with broader societal issues. For instance, the increased demand for critical minerals required for AI technologies and electric vehicles has led to disputes over indigenous territories <sup>95</sup>. As such, AI governance must also consider the rights and interests of indigenous peoples and invite them to the conversation.

### Autonomous Weapons Systems

The most contentious issue is the use of AI in warfare, particularly autonomous weapons systems. Developing and deploying such systems, often termed "killer robots," have spurred international debates and calls for regulations. Campaigns such as the "Campaign to Stop Killer Robots" petition have urged new international laws under the 1980 Convention on Conventional Weapons to address this issue <sup>96</sup>.

In summary, as we embrace AI's benefits, we must also be vigilant about its potential pitfalls and long-term concerns. From safeguarding our environment and respecting indigenous rights to preventing a new arms race, the governance of AI is as much about these broader societal issues as it is about technology itself.

# Recommendations

Drawing on the comprehensive exploration of global AI governance, this report proposes informed recommendations for the future. The following section outlines pragmatic steps aimed at harmonising AI development worldwide. These recommendations, directed towards policymakers, business leaders, and stakeholders, are derived from our study of various national and regional frameworks. The intent is to foster AI development and deployment that respects human rights, promotes societal benefits, and encourages innovation while addressing potential risks. As we present these recommendations, the ultimate objective remains - to foster global cooperation for a robust, universally beneficial AI governance framework.

## Promote Transparency

Developers and AI companies should prioritise creating transparent AI systems. Addressing the 'black box' problem will improve trust in AI systems and make it easier to identify and rectify biases and inaccuracies in AI decisions [103](#), [104](#).

## Ethics in AI Development

Businesses must incorporate ethical considerations from the inception of AI system development, which involves considering the potential societal impacts, such as job displacement, privacy concerns, and economic inequality. Businesses should be prepared to conduct regular AI ethics audits and have strategies to mitigate potential negative impacts [57](#), [102](#), [105](#).

## Collaboration across Sectors

Collaboration is key to addressing AI's challenges. Policymakers, AI developers, businesses, and civil society must work together to navigate the complex AI landscape. Collaboration efforts might involve cross-sector working groups or committees addressing AI issues [102](#).

## Invest in AI Literacy and Education

Governments and educational institutions must prioritise AI literacy, which involves promoting technical skills and teaching about the ethical, societal, and philosophical issues surrounding AI. The aim should be to create a society that can engage critically with AI, understand its potential benefits and identify and challenge its potential risks [57](#), [106](#).

## Develop Robust Security Measures:

Given the potential for AI to be used maliciously or to be vulnerable to cyberattacks, there needs to be a strong focus on developing robust AI security measures. These measures include dedicated cybersecurity teams for AI, ongoing monitoring for potential vulnerabilities, and rapid response measures to address any breaches that occur [105, 107](#).

## Address the Long-term Risks of AI:

While speculative, AI's potential long-term risks, including superintelligence development, must be taken seriously. Policymakers, scientists, and AI developers should consider these potential risks in their strategic planning and invest in research to understand and mitigate them [62](#).

## Unified International Guidelines:

Policymakers should consider creating an international AI framework that respects nations' societal and cultural differences and provides a baseline for AI development and deployment. It should be a universal obligation to ensure that the development and public offering of AI technology is governed by strict principles, design controls, and processing protocols<sup>121</sup>. This unified framework will help address alignment issues and foster a sense of global cooperation in AI governance [102](#).

Numerous organisations and institutions worldwide have attempted to establish universal rules or guidelines for AI, each contributing unique perspectives and recommendations. From prominent technology companies like Google to international bodies such as the OECD and UNESCO, these principles aim to guide AI development towards benefitting humanity, ensuring transparency, fairness, and robustness while avoiding misuse and harm.

The following offers a comparative summary of these notable guidelines, presenting the diverse set of ideas that inform the evolving global discourse on AI ethics and regulation.

### Google's AI Principles

Google's seven principles emphasise the beneficial and unbiased use of AI, incorporating privacy design principles, avoiding harm, and maintaining accountability in its AI applications. They have an additional set of four application areas they will not pursue with AI, such as technologies that cause harm or are likely to be used to infringe on human rights [109](#).

## The Asilomar AI Principles

These 23 principles cover a range of considerations, including research funding, transparency, and avoiding a competitive race without adequate safety precautions for advanced AI. The principles also advocate for the overall benefit of humanity from AI and advanced technologies [110](#).

## OpenAI's Charter

OpenAI's charter focuses on the beneficial use of Artificial General Intelligence (AGI) for all, with a strong emphasis on safety and the cooperative orientation of AGI development. They pledge to assist rather than compete with other projects close to building AGI safely [111](#).

## EU Ethics Guidelines for Trustworthy AI

This set of guidelines emphasises respect for human autonomy, prevention of harm, fairness, and explicability as the core foundational principles for AI. The guidelines also outline requirements for Trustworthy AI, such as technical robustness and safety, and accountability [112](#).

## IEEE's Ethically Aligned Design

IEEE's comprehensive set of guidelines focuses on embedding human values in the design and use of autonomous and intelligent systems, and they emphasise the importance of prioritising human well-being in the AI design process [113](#).

## The Toronto Declaration

This declaration, focusing on machine learning systems, underscores the need to protect the right to equality and non-discrimination. It stresses the requirement for AI and machine learning systems to respect and protect human rights [114](#).

## The Partnership on AI

The partnership outlines tenets that aim to ensure the safe, fair, and beneficial use of AI technologies for all humanity. The tenets discourage the harmful use of AI and emphasise the importance of cooperation and collaboration in AI research and policy [115](#).

## Montreal Declaration for a Responsible Development of AI

The declaration highlights ethical principles for AI, including respect for autonomy, consideration of well-being, privacy, solidarity, democratic participation, equity, and diversity, among others. It aims to ensure the socially responsible and sustainable development of AI [116](#).

## OECD Principles on Artificial Intelligence

OECD's five principles underscore the need for AI to promote inclusive growth, sustainable development, and well-being while advocating for transparency and robustness in AI systems [117](#).

## UNESCO's recommendation on the ethics of AI

Although still under development when writing this document, this recommendation seeks to provide an ethical framework to guide [118](#).

## Establishing an International Agency for AI:

Establishing an international body as a global, neutral, non-profit entity is imperative to ensure a unified approach to the complex web of Artificial Intelligence (AI) governance. This proposed agency would unite key stakeholders, including governments, large technology corporations, non-profits, academia, and the broader society, in a concerted effort to develop governance and technical solutions for AI.

AI's broad influence extends beyond conventional software, significantly impacting the domains of robotics and automation technologies. Recognising this vast sphere of influence underlines the need for a comprehensive, globally coordinated approach to AI governance, including the interconnected realms of robotics and automation.

The proposed agency could take cues from organisations such as the International Atomic Energy Agency (IAEA) and the UN's nuclear watchdog [108](#) to ensure the ethical and safe usage of AI and associated technologies. Acting as a global thought leader on AI and robotics, it would be responsible for developing and promoting policies, frameworks, standards, and certifications to guide their responsible development and use.

The proposed agency can also collaborate with diverse organisations, including environmental and humanitarian groups like Greenpeace, the World Wildlife Fund (WWF), Amnesty International, and the Red Cross. Such partnerships could ensure that the broader impacts of AI, including its environmental and humanitarian implications, are considered in AI governance. These collaborations would further broaden the scope and relevance of the agency, strengthening its ability to address the multitude of AI's societal impacts.

Standard-setting organisations such as the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS) can provide valuable insights and strategies. Their contributions to standards development across sectors, including robotics and automation, offer instructive models for the agency's standard-setting efforts.

## Inspiration from International Bodies

The power of international bodies to influence global policy is well-illustrated by the Intergovernmental Panel on Climate Change (IPCC) <sup>98</sup>. Established by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) in 1988, the IPCC's mission is to conduct a comprehensive review and provide recommendations concerning climate change science, its socio-economic impact, and potential response strategies.

The IPCC's work culminated in the 1992 UN Framework Convention on Climate Change, highlighting the potential of an international body to drive global policy. The proposed agency could adapt the operational structures and funding mechanisms of bodies like the IPCC and IAEA while collaborating with standard-setting bodies like the ISO, IEC, and OCEANIS. Such partnerships would position the agency to comprehensively shape the future of AI, robotics, and automation governance globally.

## Fostering Consensus through Collaboration

The proposition for an international, neutral, non-profit agency focusing on the governance of AI is gaining traction. Drawing lessons from established international bodies such as the IAEA, the IPCC and standards-setting organisations like the ISO and the IEC would be beneficial as it would foster collaboration amongst diverse stakeholders <sup>99</sup>. Additionally, forging partnerships with humanitarian and environmental organisations would guarantee the consideration of the wider implications of AI in its governance. To keep such an organisation neutral and non-bias, a dedicated trust fund, pooling resources from industry stakeholders, governments, and non-profits globally, could provide the necessary financial backbone.

Emulating the inclusive processes of these organisations, the agency would engage a broad spectrum of experts who collectively volunteer their expertise to create its policies and frameworks <sup>100</sup>. This collaborative spirit ensures an inclusive approach to AI, robotics, and automation governance, capturing many perspectives, including those from environmental and humanitarian sectors.



## Addressing Criticisms

International bodies striving to provide global thought leadership by developing comprehensive policies, frameworks, standards, and certifications face regular criticism. Critics often argue about the IAEA's conflicting mandate and lack of enforcement capabilities, the perceived politicisation of the UN's nuclear watchdog, and the potential for underestimating impacts or diluting findings in the IPCC's consensus-driven approach <sup>101</sup>. Criticisms aimed at ISO and IEC commonly revolve around accessibility, the pace of normal development, and power imbalances.

The agency must acknowledge and address these criticisms while learning from these models. Heeding these criticisms, the proposed agency must foster consensus while maintaining robustness in its governance approach, particularly concerning AI, robotics, and automation. The agency's agility and adaptability will be paramount to effectively navigate the rapidly evolving landscape of technology, ensuring the development and application of AI and related technologies serve as tools of equitable advancement and societal good.

# Conclusion

The growth and reach of Artificial Intelligence technology over the past decade have significantly impacted numerous aspects of society, the economy, and the environment. AI, robotics, and automation have transcended national boundaries and sectors, becoming integral to the global socio-economic fabric. Dissimilar regulatory frameworks, cultural variations, and the requirement for universal standards and definitions currently define the governance landscape.

Despite the various governance frameworks in place, AI's rapid expansion has also brought about issues concerning rights, safety, ethics, and more. This fragmentation poses significant challenges to ensuring AI's safe, equitable, and beneficial development and deployment [93, 94](#). The potential of this technology, the myriad risks, implications, and complexities accompanying the continued advancement of AI, along with associated fields such as robotics and automation, demand a comprehensive global approach and necessitate the establishment of an International Agency capable of tackling these issues effectively and responsibly. [3, 93](#).

In conclusion, the urgent creation of an International Agency for AI is not merely a desirable proposition but an indispensable necessity. In light of Wyatt Tessari's profound statement in his recent op-ed, 'Currently, AI systems are less capable than humans in many key areas. But that won't hold true for long.' [122](#) the need for immediate and decisive action becomes unmistakably clear. Establishing an International Agency for AI would provide a unique opportunity to offer unified global thought leadership on various critical issues. These include the development of international standards, certification processes, regulatory frameworks, and ethical guidelines.

Drawing inspiration from successful models like the IPCC and addressing their shortcomings, we can pave the way for an effective, inclusive, and responsible international AI governance structure. By fostering international collaboration, it could harmonise various national and regional approaches and facilitate consensus-building on contentious issues such as privacy norms, algorithmic bias, and the use of AI in sensitive domains such as security, healthcare, and employment [78, 95](#).

Furthermore, the agency could serve as a neutral platform for dialogue and cooperation among diverse stakeholders, including governments, industry, academia, and civil society. Promoting knowledge sharing, capacity building, and stakeholder engagement could ensure everyone has a voice in the AI governance discourse [96](#).

Finally, the agency could play a critical role in monitoring the rapid technological advancements in AI, robotics, and automation and their societal implications. It could anticipate potential risks, identify regulatory gaps, and propose proactive and timely interventions [97](#).

As we enter the era of the “Fourth Industrial Revolution,” where AI and associated technologies are poised to redefine society, economy, and human life, it is imperative to move forward with foresight, unity, and prudence. An International Agency for AI, underpinned by a commitment to transparency, inclusivity, and the public interest, could provide the leadership and guidance needed to navigate this uncharted terrain. By steering the wheel of AI governance, we can ensure that these technologies are harnessed for the greater good of humanity <sup>48</sup>.

# References

1. "Declaration of Cooperation on Artificial Intelligence." European Commission, 2018. (<https://ec.europa.eu/newsroom/dae/redirection/document/50951>)
2. "Coordinated Plan on Artificial Intelligence." European Commission, 2018. ([https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=56017](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56017))
3. "Artificial Intelligence Act." European Commission, 2021. ([https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF))
4. "Ethics Guidelines for Trustworthy AI." High-Level Expert Group on Artificial Intelligence, European Commission, 2019. ([https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=60419](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419))
5. "A European Strategy for Data." European Commission, 2020. (<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066>)
6. Vestager, M. (2021). "Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence" European Commission, 2021. ([https://ec.europa.eu/commission/presscorner/detail/en/IP\\_21\\_1682](https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682))
7. "Establishing a pro-innovation approach to regulating AI" UK Government, 2023." ([https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1092630/CP\\_728\\_-\\_Establishing\\_a\\_pro-innovation\\_approach\\_to\\_regulating\\_AI.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1092630/CP_728_-_Establishing_a_pro-innovation_approach_to_regulating_AI.pdf))
8. "Equality Act 2010." UK Legislation, 2010 (<https://www.legislation.gov.uk/ukpga/2010/15/contents>)
9. "Human Rights Act 1998." UK Legislation, 1998 (<https://www.legislation.gov.uk/ukpga/1998/42/contents>)
10. "AI Council's AI Roadmap" UK Government, 2021." ([https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/949539/AI\\_Council\\_AI\\_Roadmap.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949539/AI_Council_AI_Roadmap.pdf))
11. Shadbolt, Nigel. "Architectures for Autonomy: Towards an Equitable Web of Data in the Age of AI" WWW '20: Proceedings of The Web Conference 2020 (<https://dl.acm.org/doi/10.1145/3366423.3382668>)
12. Smith, M., & Neupane, S., (2021)" Canada's artificial intelligence landscape: The perspective of state and market actors." Technology in Society, 64, 101465
13. "Bill C-427: An Act to amend the Department of Agriculture and Agri-Food Act (excellence in agricultural innovation)." Parliament of Canada, 2019 (<https://www.parl.ca/LegisInfo/en/bill/42-1/c-427>)
14. "Bill C-27: An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act." Parliament of Canada, 2023 (<https://www.parl.ca/LegisInfo/en/bill/44-1/c-27>)
15. Treasury Board of Canada Secretariat, Directive on Automated Decision-Making, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>
16. "Bill C-26: An Act Respecting Cyber Security" Parliament of Canada, 2023 (<https://parl.ca/DocumentViewer/en/44-1/bill/C-26/first-reading>)
17. "Bill C-43: An Act respecting a payment to be made out of the Consolidated Revenue Fund to support a pan-Canadian artificial intelligence strategy." Parliament of Canada, 2017 (<https://www.parl.ca/LegisInfo/en/bill/42-1/c-43>)
18. "AI guru Yoshua Bengio says regulation too slow, warns of 'existential' threats" The Canadian Press, May 24, 2023 ([https://www.thecanadianpressnews.ca/politics/ai-guru-yoshua-bengio-says-regulation-too-slow-warns-of-existential-threats/article\\_fa558989-35e2-5051-a7c4-5243e4982414.html](https://www.thecanadianpressnews.ca/politics/ai-guru-yoshua-bengio-says-regulation-too-slow-warns-of-existential-threats/article_fa558989-35e2-5051-a7c4-5243e4982414.html))
19. "Yoshua Bengio Weighs in on the Pause and Building a World Model" Hackernoon, April 2023 (<https://hackernoon.com/yoshua-bengio-weighs-in-on-the-pause-and-building-a-world-model/>)
20. "AI Training Act (S.2551)" 117th Congress, (2021-2022) (<https://www.congress.gov/bill/117th-congress/senate-bill/2551>)
21. "Infrastructure Investment and Jobs Act (H.R.3684)" 117th Congress, (2021-2022) (<https://www.congress.gov/bill/117th-congress/house-bill/3684>)
22. "Automated Decision Systems Accountability Act" California Legislature, 2020 ([https://leginfo.ca.gov/faces/billTextClient.xhtml?bill\\_id=201920200AB2004](https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB2004))

23. Senate Commerce Committee Hearing on Artificial Intelligence, May 2023  
(<https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence>)
24. "Artificial Intelligence and the Bill of Rights for an AI-Powered World." White House OSTP, 2023  
(<https://www.whitehouse.gov/ostp/>)
25. "AI Risk Management Framework." NIST, 2023  
(<https://www.nist.gov/topics/artificial-intelligence/ai-risk-management-framework>)
26. Fei-Fei Li, "Introduction to Stanford HAI: Fei-Fei Li" Stanford HAI (2019)  
(<https://www.youtube.com/watch?v=XnhfeNDc0el>)
27. "State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan," State Council of the People's Republic of China, 2017.  
(<https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>)
28. Kania E., Triolo P., Webster G., (2018). "Translation: Chinese government outlines AI ambitions through 2020" DigiChina 2018  
([https://www.iaist.ac.jp/~bao/AI/OtherAIstrategies/Translation %20Chinese%20government%20outlines%20AI%20ambitions%20through%202020.pdf](https://www.iaist.ac.jp/~bao/AI/OtherAIstrategies/Translation%20Chinese%20government%20outlines%20AI%20ambitions%20through%202020.pdf))
29. "Understanding China's AI Strategy" Center for a New American Security. 2019  
(<https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy>)
30. "Chinese Views of Big Data Analytics" RAND Corporation, 2020.  
([https://www.rand.org/content/dam/rand/pubs/research\\_reports/RRA100/RRA176-1/RAND\\_RRA176-1.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RRA100/RRA176-1/RAND_RRA176-1.pdf))
31. "Beijing AI Principles," Beijing Academy of Artificial Intelligence, 2019  
(<https://link.springer.com/content/pdf/10.1007/s11623-019-1183-6.pdf>)
32. Kai-Fu Lee, "AI Superpowers: China, Silicon Valley, and the New World Order," Houghton Mifflin Harcourt, 2018  
([https://www.google.ca/books/edition/AI\\_Superpowers/Xb9wDwAAQBAJ?hl=en&qbpv=1&dq=Ai+Superpowers:+China,+Silicon+Valley,+and+the+New+World&printsec=frontcover](https://www.google.ca/books/edition/AI_Superpowers/Xb9wDwAAQBAJ?hl=en&qbpv=1&dq=Ai+Superpowers:+China,+Silicon+Valley,+and+the+New+World&printsec=frontcover))
33. "Model AI Governance Framework," Infocomm Media Development Authority of Singapore, 2019  
(<https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/smodelaigovframework2.pdf>)
34. "Advisory Council on the Ethical Use of AI and Data," Infocomm Media Development Authority of Singapore.  
(<https://www.imda.gov.sg/resources/press-releases-and-factsheets/archived/imda/press-releases/2018/comp-osition-of-the-advisory-council-on-the-ethical-use-of-ai-and-data>)
35. "Guide to Data Protection by Design for ICT Systems," Personal Data Protection Commission, Singapore, 2021.  
(<https://www.pdpc.gov.sg/help-and-resources/2019/05/guide-to-data-protection-by-design-for-ict-systems>)
36. "How Singapore intends to harness AI for the public good" The Straits Times, 2023.  
(<https://www.straitstimes.com/opinion/how-singapore-intends-to-harness-ai-for-the-public-good>)
37. "Securing Singapore's Future in the Digital Age" Singapore Ministry of Communications and Information, 2019.  
(<https://www.mci.gov.sg/pressroom/news-and-stories/pressroom/2019/11/keynote-address-by-mr-s-iswara-n-at-the-asia-house-global-trade-dialogue-on-7-nov-2019>)
38. India's National Strategy for AI:  
(<https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf>)
39. Lee, S. K., & Trivedi, S. (2019). "AI Policy – India." The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.  
([https://www.researchgate.net/publication/332458783\\_The\\_IEEE\\_Global\\_Initiative\\_on\\_Ethics\\_of\\_Autonomous\\_and\\_Intelligent\\_Systems](https://www.researchgate.net/publication/332458783_The_IEEE_Global_Initiative_on_Ethics_of_Autonomous_and_Intelligent_Systems))
40. "India's Artificial Intelligence Revolution" India Brand Equity Foundation, 2022.  
(<https://www.ibef.org/download/AI-Revolution.pdf>)
41. "How Agritech Startups Are Leveraging AI & IoT To Reinvent Agricultural Sector" Inc42 Media, 2022  
(<https://inc42.com/resources/how-agritech-startups-are-leveraging-ai-iot-to-reinvent-agricultural-sector>)
42. "Top Data Science & AI Courses That Were Introduced In 2020 In India" Analytics India Magazine, 2020  
(<https://analyticsindiamag.com/top-data-science-ai-courses-that-were-introduced-in-2020-in-india>)

43. "France focuses on digital economy, AI, space to boost ties with India" HindustanTimes, 2019. (<https://www.hindustantimes.com/india-news/can-work-with-india-on-defence-ai-and-cyber-security-france-officials/story-2DdyJZX85yXH3hWFw92ZZI.html>)
44. "Maintaining American Leadership in Artificial Intelligence" US Government, Executive Order 2019 (<https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence>)
45. "AI Sector Deal" UK Government, Policy Paper 2019 (<https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>)
46. "Data Protection Act" UK Government 2018 (<https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>)
47. "Coordinated Plan on Artificial Intelligence" EU Parliament 2018 (<https://digital-strategy.ec.europa.eu/en/policies/plan-a>)
48. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics. (<https://www.nature.com/articles/s42256-019-0088-2>)
49. "Canada's Pan-Canadian AI Strategy" Government of Canada 2022 (<https://ised-isde.canada.ca/site/ai-strategy/en>)
50. "Personal Information Protection and Electronic Documents Act" Government of Canada (<https://laws-lois.justice.gc.ca/PDF/P-8.6.pdf>)
51. Singapore's National AI Strategy: (<https://www.smartnation.gov.sg/initiatives/artificial-intelligence>)
52. "Cybersecurity Law of the People's Republic of China" PRC 2017 (<https://digichina.stanford.edu/work/translation-cybersecurity-law-of-the-peoples-republic-of-china-effective-june-1-2017>)
53. "Personal Data Protection Act 2012" Singapore Government 2012 (<https://sso.agc.gov.sg/Act/PDPA2012>)
54. Papernot, N., et al. (2016). The limitations of deep learning in adversarial settings. In Proceedings of the 1st IEEE European Symposium on Security and Privacy. (<https://ieeexplore.ieee.org/document/7467366>)
55. Amodei, D., et al. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565. (<https://arxiv.org/abs/1606.06565>)
56. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608 (<https://arxiv.org/abs/1702.08608>)
57. Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company ([https://www.google.ca/books/edition/The\\_Second\\_Machine\\_Age\\_Work\\_Progress\\_and/WiKwAqAAQBAJ](https://www.google.ca/books/edition/The_Second_Machine_Age_Work_Progress_and/WiKwAqAAQBAJ))
58. Pasquale, F. (2015). The black box society. Harvard University Press (<https://www.hup.harvard.edu/catalog.php?isbn=9780674970847>)
59. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104, 671 (<https://www.courts.ca.gov/documents/BTB24-21-2.pdf>)
60. Bessen, J. E. (2019). AI and Jobs: The Role of Demand. NBER Working Paper No. 24235 ([https://www.nber.org/system/files/working\\_papers/w24235/w24235.pdf](https://www.nber.org/system/files/working_papers/w24235/w24235.pdf))
61. Brundage, M., et al. (2018). The malicious use of Artificial Intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228 (<https://arxiv.org/abs/1802.07228>)
62. Bostrom, N. (2014). Superintelligence. Paths, dangers, strategies. Oxford University Press (<https://global.oup.com/academic/product/superintelligence-9780199678112>)
63. Sharif, M. et al. (2016). Accessorise to a crime: Real and stealthy attacks on state-of-the-art face recognition. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (<https://dl.acm.org/doi/10.1145/2976749.2978392>)
64. Anderson, H., & Anderson, P. (2018). AI and the Future of Cybersecurity. ITNOW, 60(1), 34-37
65. Harari, Y. N. (2018). 21 lessons for the 21st century. Random House (<https://www.penguinrandomhouse.com/books/579741/21-lessons-for-the-21st-century-by-yuval-noah-harari>)
66. Duggan, D., & Carlin, S. (2007). An adaptive approach to network intrusion detection. In Proceedings of the 13th Communications & Networking Symposium (<https://dl.acm.org/doi/proceedings/10.1145/1287853>)
67. Russell, S. (2015). Research priorities for robust and beneficial artificial Intelligence. AI Magazine, 36(4), 105-114 ([https://futureoflife.org/data/documents/research\\_priorities.pdf](https://futureoflife.org/data/documents/research_priorities.pdf))

68. Vincent, J. (2018). Google is building AI to analyse drone footage, and Pentagon is interested. The Verge (<https://www.theverge.com/2018/3/6/17086276/google-ai-military-drone-analysis-pentagon-project-maven-tensorflow>)
69. He, D. et al. (2017). A predictive maintenance model using contextual deep learning. In Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining (<https://dl.acm.org/doi/proceedings/10.1145/3097983?tocHeading=heading2>)
70. Clark, S., & Levy, D. (2018). Cybersecurity, Artificial Intelligence and the scale problem. In Proceedings of the 2018 Conference on Artificial Intelligence, Ethics, and Society (<https://dl.acm.org/doi/proceedings/10.1145/3278721>)
71. Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. WW Norton & Company (<https://wwnorton.com/books/Army-of-None/>)
72. Glaessgen, E., & Stargel, D. (2012). The digital twin paradigm for future NASA and US Air Force vehicles. In Proceedings of the 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference (<https://arc.aiaa.org/doi/book/10.2514/MSDM12>)
73. Kott, A., et al. (2018). AI and information warfare: The new mental battlefield. Computing Edge, 7(8), 24-31 (<https://s3.amazonaws.com/media.hudson.org/Hudson%20Transcript%20-%20The%20Future%20of%20Warfare.pdf>)
74. Segler, M. H., et al. (2018). Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. ACS Central Science (<https://pubs.acs.org/doi/full/10.1021/acscentsci.7b00512>)
75. Gibson, D. G., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesised genome. Science, 329(5987), 52-56 (<https://www.science.org/doi/10.1126/science.1190719>)
76. Zhavoronkov, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology, 37(9), 1038-1040 (<https://www.nature.com/articles/s41587-019-0224-x>)
77. Li, X., et al. (2018). Optimising Chemical Reactions with Deep Reinforcement Learning. ACS Central Science (<https://pubs.acs.org/doi/10.1021/acscentsci.7b00492>)
78. Russell, S. J. (2019). "Human Compatible: Artificial Intelligence and the Problem of Control." Viking. (<https://www.penguinrandomhouse.com/books/566677/human-compatible-by-stuart-russell/>)
79. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. (2018). (<https://www.repository.cam.ac.uk/items/d654418d-1c12-4024-85d5-ccd614c32ef3>)
80. Will robots and AI take your job? The economic and political consequences of automation (2018). Brookings Institution. (<https://www.brookings.edu/blog/techtank/2018/04/18/will-robots-and-ai-take-your-job-the-economic-and-political-consequences-of-automation/>)
81. Artificial Intelligence and Legal Ethics: Understanding the Future Challenges & Opportunities. (2021). American Bar Association. (<https://www.americanbar.org/events-cle/ecd/ondemand/413900125/>)
82. Tegmark, M (2017) "Life 3.0" Penguin Random House. (<https://www.penguinrandomhouse.com/books/530584/life-30-by-max-tegmark/>)
83. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. (<https://aclanthology.org/P19-1355/>)
84. Whittlestone, Jess et al. "Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research." Nuffield Foundation, 2019. (<https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf>)
85. "Global AI in Financial Services Survey", Cambridge Centre for Alternative Finance, 2020. ([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3532038](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3532038))
86. "Preparing for the Future of Artificial Intelligence", Executive Office of the President National Science and Technology Council Committee on Technology, 2016. ([https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf))
87. "The Global AI Agenda: Promise, Reality, and a Future of Data Sharing", MIT Technology Review, 2020. (<https://mittrinsights.s3.amazonaws.com/Alagenda2020/GlobalAlagenda.pdf>)
88. "Ethics of Artificial Intelligence and Robotics", Stanford Encyclopedia of Philosophy, 2020. (<https://plato.stanford.edu/entries/ethics-ai/>)



89. Floridi, Luciano. "The logic of information: a theory of philosophy as conceptual design." Oxford University Press, 2019. (<https://academic.oup.com/book/27824>)
90. Dehaene, Stanislas et al. "What is consciousness, and could machines have it?" Science, 2017. (<https://pubmed.ncbi.nlm.nih.gov/29074769/>)
91. Bryson, Joanna J. "Artificial intelligence and pro-social behaviour." Springer, 2015. ([https://www.researchgate.net/profile/Joanna-Bryson/publication/292354936\\_Artificial\\_Intelligence\\_and\\_Pro-Social\\_Behaviour/links/5772b94508ae2b93e1a7cbf7/Artificial-Intelligence-and-Pro-Social-Behaviour.pdf](https://www.researchgate.net/profile/Joanna-Bryson/publication/292354936_Artificial_Intelligence_and_Pro-Social_Behaviour/links/5772b94508ae2b93e1a7cbf7/Artificial-Intelligence-and-Pro-Social-Behaviour.pdf))
92. Legg, Shane and Hutter, Marcus. "A Collection of Definitions of Intelligence." Frontiers in Artificial Intelligence and Applications, Vol.157 (2007) 17-24. (<https://arxiv.org/pdf/0706.3639.pdf>)
93. Lin, P., Abney, K., & Bekey, G. A. (2012). Robot ethics: the ethical and social implications of robotics. MIT Press. (<https://mitpress.mit.edu/9780262526005/robot-ethics/>)
94. Mulligan, D. K., & Bamberger, K. A. (2018). Saving governance-by-design. Cal. L. Rev., 106, 697. (<https://lawcat.berkeley.edu/record/1128572?ln=en>)
95. Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial Intelligence. Cambridge Handbook of Artificial Intelligence, 316-334. (<https://nickbostrom.com/ethics/artificial-intelligence.pdf>)
96. Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines, 30(1), 99-120. (<https://link.springer.com/article/10.1007/s11023-020-09517-8>)
97. Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning. In Proceedings of the 52nd Hawaii International Conference on System Sciences. (<http://pdfs.semanticscholar.org/625d/a63503d70be79cf5a6454686baae4a9256fa.pdf>)
98. "MOU between UNEP and WMO on IPCC" IPCC. (1988). ([https://www.ipcc.ch/site/assets/uploads/2019/06/MOU\\_between\\_UNEP\\_and\\_WMO\\_on\\_IPCC-1989.pdf](https://www.ipcc.ch/site/assets/uploads/2019/06/MOU_between_UNEP_and_WMO_on_IPCC-1989.pdf))
99. "IPCC Principles Governing IPCC Work." IPCC. (1998). (<https://www.ipcc.ch/site/assets/uploads/2018/09/ipcc-principles.pdf>)
100. "Structure of the IPCC" IPCC. (2023). (<https://www.ipcc.ch/about/structure/>)
101. Hulme, M., & Mahony, M. (2010). Climate change: What do we know about the IPCC? Progress in Physical Geography, 34(5), 705-718. ([https://www.researchgate.net/profile/Mike-Hulme/publication/215646136\\_Climate\\_change\\_What\\_do\\_we\\_know\\_about\\_the\\_IPCC/links/00b4952c5959f0785a000000/Climate-change-What-do-we-know-about-the-IPCC.pdf](https://www.researchgate.net/profile/Mike-Hulme/publication/215646136_Climate_change_What_do_we_know_about_the_IPCC/links/00b4952c5959f0785a000000/Climate-change-What-do-we-know-about-the-IPCC.pdf))
102. Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689-707. (<https://link.springer.com/article/10.1007/s11023-018-9482-5>)
103. Castelvechi, D. (2016). Can we open the black box of AI? Nature News, 538(7623), 20-23. ([https://www.nature.com/news/polopoly\\_fs/1.20731!/menu/main/topColumns/topLeftColumn/pdf/538020a.pdf](https://www.nature.com/news/polopoly_fs/1.20731!/menu/main/topColumns/topLeftColumn/pdf/538020a.pdf))
104. 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 (<https://www.nature.com/articles/s42256-019-0048-x>)
105. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. (<https://www.sciencedirect.com/science/article/abs/pii/S1566253519308103>)
106. Rainie, L., & Anderson, J. (2017). The future of jobs and jobs training. Pew Research Center, 3. (<https://www.pewresearch.org/internet/2017/05/03/the-future-of-jobs-and-jobs-training/>)
107. Taddeo, M., & Floridi, L. (2018). How AI can be a force for good: An ethical framework will help to harness the potential of AI while keeping humans in control. Science (<https://www.science.org/doi/10.1126/science.aat5991>)
108. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': the US, EU, and UK approach. Science and engineering ethics, 24(2), 505-528. (<https://link.springer.com/article/10.1007/s11948-017-9901-7>)
109. "Google AI Principles" Google (<https://ai.google/principles/>)



110. "The Asilomar AI Principles" FLI (<https://futureoflife.org/ai-principles/>)
111. "OpenAI Charter" OpenAI, (<https://openai.com/chart/m>)
112. "Ethics guidelines for trustworthy AI" European Commission (<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>)
113. "The IEEE Global Initiative "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (<https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>)
114. "Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems" Access Now, 2018 ([https://www.accessnow.org/wp-content/uploads/2018/08/The-Toronto-Declaration\\_ENG\\_08-2018.pdf](https://www.accessnow.org/wp-content/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf))
115. Partnership on AI (<https://partnershiponai.org/about/>)
116. "Montreal Declaration for a Responsible Development of AI" Université de Montréal (<https://www.montrealdeclaration-responsibleai.com/the-declaration>)
117. "OECD AI Principles "Organisation for Economic Cooperation and Development (<https://oecd.ai/en/ai-principles>)
118. "UNESCO Recommendation on the Ethics of Artificial Intelligence" UNESCO 2021 (<https://www.ohchr.org/sites/default/files/2022-03/UNESCO.pdf>)
119. "Emergency Resolution: Effective Regulation Of AI Technology [For Democracy, Sustainability And Social Good]" Global Green 2023 ([https://globalgreens.org/gg\\_resolution/emergency-resolution-effective-regulation-of-ai-technology-for-democracy-sustainability-and-social-good/](https://globalgreens.org/gg_resolution/emergency-resolution-effective-regulation-of-ai-technology-for-democracy-sustainability-and-social-good/))
120. Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Conference on Fairness, Accountability, and Transparency (FAccT' 21), March 3–10, 2021, Virtual Event, Canada. ACM (<https://doi.org/10.1145/3442188.3445922>)
121. Christine Elwell "Artificial General Intelligence – Regulating to Promote Human Control" Green Party of Canada 2023 ([https://drive.google.com/file/d/15Tuwf-oZgDg86bkkltlHAIbXacc8Q/view?usp=share\\_link](https://drive.google.com/file/d/15Tuwf-oZgDg86bkkltlHAIbXacc8Q/view?usp=share_link) )
122. Wyatt Tessari L'Allié "AI a bigger threat to humanity than climate change or pandemics" Toronto Star 2023 (<https://www.thestar.com/opinion/contributors/2023/05/27/ai-a-bigger-a-threat-to-humanity-than-climate-change-or-pandemics.html>)