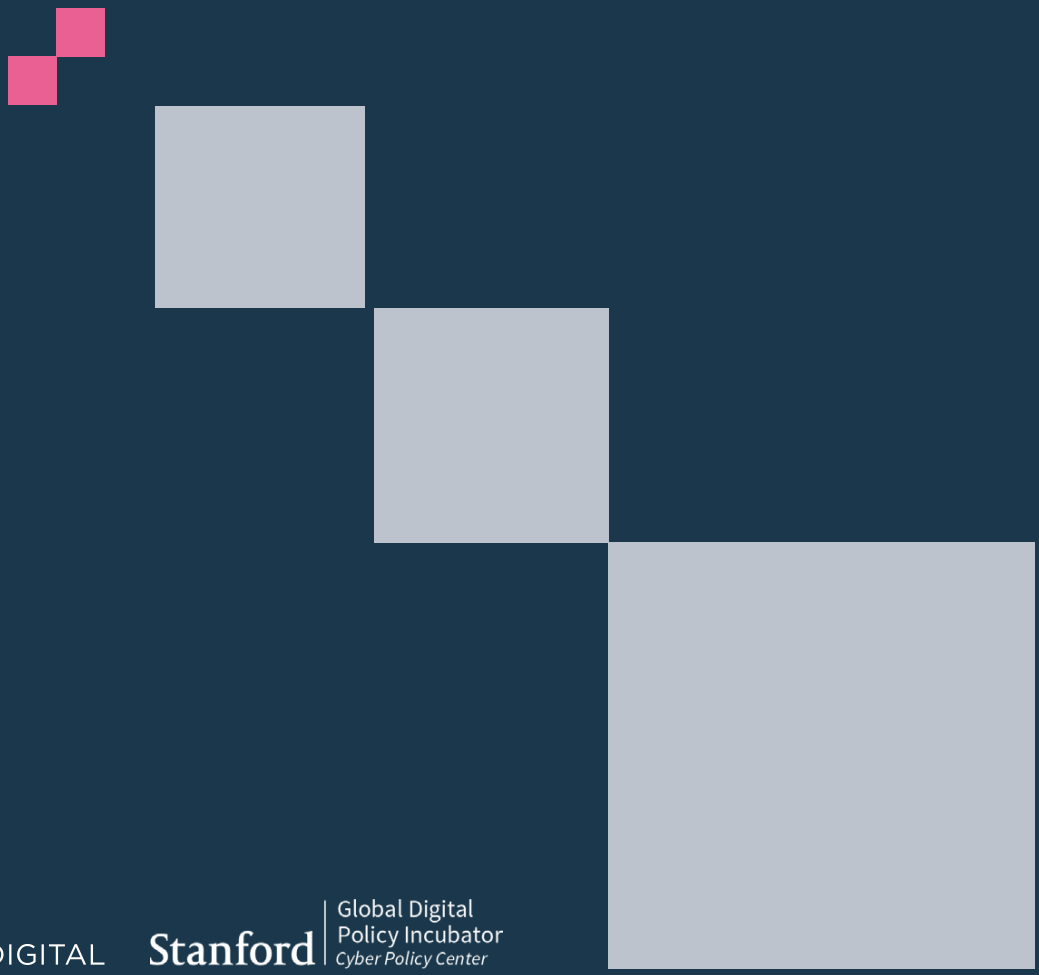
National Artificial Intelligence Strategies and Human Rights: A Review

Second edition
April 2021



GLOBAL PARTNERS DIGITAL

Stanford | Global Digital Policy Incubator
Cyber Policy Center

Contents

Acknowledgments

This report was authored by Charles Bradley and Richard Wingfield at Global Partners Digital and Megan Metzger at the Global Digital Policy Incubator at the Stanford Cyber Policy Center, with research assistance from Madeline Libbey and Amy MacKinnon.

The authors would like to thank those who generously donated their time through interviews and providing invaluable feedback on an early draft of this report: Danya Centeno, Philip Dawson, Gallit Dobner, Eileen Donahoe, Laura Okkonen, Natalie Evans Harris, Francesca Fanucci, Steven Feldstein, Stefan Heumann, Fanny Hidvegi, Patrik Hiselius, Katharina Höne, Zach Lampell, Grace Mutung'u, Cailean Osborne, Philippe-André Rodriguez, Vidushi Sharma, Lisa Vermeer and Marlena Wisniak.

Executive Summary

A National Artificial Intelligence Strategy (NAS) is a document, ordinarily developed by a government, which sets out its broad, strategic approach to artificial intelligence ("AI), including specific areas of focus and activities they will undertake which relate to AI. In doing so, an NAS attempts to coordinate government policies in order to maximize the potential benefits for the economy and society, while minimizing the potential costs. Since 2017, over 40 states and regional intergovernmental organizations have published them, with many more in the process of development.

The purpose of this report is to understand how human rights have (or have not) been incorporated into NASs up to this point, to identify emerging trends at the regional level, and to provide recommendations for how to incorporate them in the future. The recent uptick in the number of NASs reflects the fact that governments in many states are beginning to think seriously about how they will harness the economic and social benefits offered by AI and remain competitive in the global market for this technology. It is critical that as states develop their strategies, they also build an approach that ensures that this innovation does not come at the cost of human rights violations.

The importance of human rights in National Artificial Intelligence Strategies

States have an obligation under international human rights law to protect the human rights of all people within their territory and jurisdiction from violations, caused either by their own policies or practices, or the acts of third parties, such as businesses or other individuals. Human rights considerations are applicable to all areas of government policy and practice, including AI policy and AI applications used by governments and other entities in society.

NASs represent a government's (or group of governments') comprehensive roadmap for how they intend to approach AI. This includes how AI will be regulated, how they will support innovation in the AI sector, and how they will address the impact on people's lives and on labor. As the implementation of strategies to improve AI training or to develop new technologies moves forward, it may do so without fully addressing risks related to the violation of human rights. It is therefore critical that NASs set out how the protection of human rights will be ensured.

Despite the critical importance of considering human rights when it comes to AI policy, to date relatively few NASs deeply engage with the human rights impacts of this technology. For governments in some states, this may be because they simply do not place a high priority on the protection of human rights in their policymaking in general. Some governments that do prioritize human rights may nonetheless find it challenging to outline approaches to their protection when it stands in contrast to other goals the government has for AI, such as enhanced economic or geopolitical competitiveness. Still others may simply not know what it would look like to create an NAS that is rights-respecting in this policy area. States' obligations to protect human rights do not go away, however, because they are unclear or inconvenient.

Some have suggested new ethical frameworks for the governance of AI. In some cases, this is an attempt to circumvent the human rights frameworks entirely, or the parts of them that governments find inconvenient. In other cases, it is an attempt to go beyond the human rights frameworks and be even more protective. It is critical to note that nothing prevents governments from going beyond what is protected by human rights frameworks in their governance of AI. There are, however, persuasive reasons to use the existing human rights

framework as the foundation for the regulation of AI. That is, human rights should be the floor on which any other governance approach builds. Some key reasons for this include:

- The international human rights framework centers the human person and makes them the focal point of governance.
- The international human rights framework includes provisions that address the most pressing societal concerns about AI.
- The international human rights framework establishes and clearly defines the roles and responsibilities of both governments and the private sector which is critical in the context of a technology which will require oversight from both governments and private companies.
- Finally, although interpreted and implemented in different ways around the world, the international human rights framework enjoys a level of geopolitical recognition and status under international law that any newly emergent ethical framework would be unlikely to match.

Analysis of whether and how human rights are currently incorporated into NASs

To understand whether and how NASs currently incorporate human rights, we analyzed every strategy that had been formally adopted by a state before 1 March 2021. Our analysis found that:

- A majority of the strategies made explicit reference to human rights in their text. However, the depth of subsequent engagement in the human rights framework and its application to AI varied considerably with most of these strategies including human rights only in passing.
- Many strategies referenced the need for ethical frameworks, ethical approaches, or human-centered approaches to AI. At times, there was some overlap here with a human rights-based approach, but often a lack of clarity in how this would be implemented or even in defining what these frameworks or approaches would look like.
- A number of strategies referenced specific human rights particularly impacted by AI, suggesting prioritization. The right to privacy was the most commonly mentioned, followed by the right to equality / non-discrimination.
- Some strategies also engaged with human rights issues, without specifically referencing human rights. For example, a number of strategies included substantial analyses of the implications of AI on the future of work, thereby engaging in issues relating to the right to work, but without directly addressing the human rights dimensions of the question.
- In all but a very small number of cases, there was a lack of depth and specificity on how human rights should be protected. While almost all strategies called for the mitigation of potential harms (either by ensuring the protection of human rights or by using an alternative (e.g. ethical) approach), strategies largely failed to set out any specific details of how this should be done in practice. This absence of detail stood in stark contrast to other parts of these strategies which were often quite specific and detailed.
- Interviews conducted for this project also highlighted the fact that the text of a strategy is only part of the story. A critical component of how human rights are (or are not) protected when it comes to AI in a particular state is the implementation of the strategy. Without clear and specific commitments that will ensure that human rights are protected in practice, or the creation of incentives or institutions that will promote such protections, even the strongest language related to human rights in an NAS will only be words. In the most extreme cases, language around human rights in an NAS may even be included by governments who have little desire to ensure that they are protected in practice, but seek to legitimize their strategy externally and among domestic stakeholders.

Recommendations for incorporating human rights into NASs

Based on the current state of NASs globally, there are a number of suggestions for steps governments can take to ensure that their NAS lays the groundwork for a human rights-respecting approach to AI policy. These are outlined in more detail in the assessment tool that is associated with this report (Annex 1), but they fall into a few core categories.

- **Include human rights explicitly and throughout the strategy**: Thinking about the impact of AI on human rights, and how to mitigate the risks associated with those impacts should be core to an NAS. Each section should consider the risks and opportunities AI provides as they relate to human rights, with a specific focus on at-risk, vulnerable and marginalized communities.
- **Outline specific steps to be taken to ensure human rights are protected**: As strategies engage with human rights, they should include specific goals, commitments or actions to ensure that human rights are protected.
- **Build in incentives or specific requirements to ensure rights-respecting practice:** Governments should take steps within their strategies to incentivize human rights-respecting practices and actions across all sectors, as well as to ensure that their goals with regards to the protection of human rights are fulfilled.
- **Set out grievance and remediation processes for human rights violations:** An NAS should look at the existing grievance and remedial processes available for victims of human rights violations relating to AI and determine whether they are sufficient. These processes (including their legislative underpinnings) may need revision in light of the particular nature of AI as a technology, or capacity-building of those involved in these processes so that they are able to receive complaints which involve AI.
- **Recognize the regional and international dimensions to AI policy:** NASs should clearly identify relevant regional and global fora and processes relating to AI, and the means by which the government will promote human rights-respecting approaches and outcomes through proactive engagement in those processes.
- **Include human rights experts and other stakeholders in the drafting of NASs:** When drafting an NAS, the government should ensure that experts on human rights and the impact of AI on human rights are a core part of the drafting process. These should include not only general human rights organizations, but also a broad range of civil society organizations and other stakeholders representing communities that may be adversely affected by AI or benefit particularly from certain applications.

Regional developments and trends in Artificial Intelligence governance

As the number of published NASs increases, trends at the regional and sub-regional level are emerging. These trends relate not only to the content of those NASs but also their framing and the extent to which human rights are (or are not) considered. At the same time, while NASs provide an opportunity for a government to set out a holistic approach towards AI, including its governance, they are not the only means through which a government can steer the direction of AI. Many governments have taken actions or measures in relation to AI which sufficiently influence the development, use and governance of AI such that human rights can - and should - be considered in a way comparable to that outlined in the earlier sections of this report.

- In **North America and the Caribbean**, both Canada and the United States have adopted NASs. In the absence of a formal governmental strategy, the government of Mexico has nonetheless taken action to steer the development and use of AI within government, publishing "Principles and Guidance for Impact Analyses for the Development and Use of Systems Based on Artificial Intelligence In the Federal Public Administration" in 2018, which include a number of human rights considerations. No Caribbean country has adopted a NAS, nor have there been any commitments by governments in the region to develop one. However, the Caribbean Artificial Intelligence Initiative, established in 2020 under the auspices of the UNESCO Cluster Office for the Caribbean has the objective of developing "a sub-regional strategy on the ethical use of AI.
- Of the **Central and South American** countries, only three have published NASs: Argentina, Colombia and Uruguay. All three explicitly emphasized the importance of ensuring respect for human rights throughout the text. A number of other governments have committed to developing a strategy, including Brazil and Chile. There are also initiatives taking place in other parts of the region. Costa Rica, for example, is considering AI through a High-Level Commission for Digital Government of the Bicentennial.
- A total of 24 **European** countries have adopted NASs and many others are in the process of developing one, in part because of the EU's push in its Coordinated Plan on Artificial Intelligence. European countries have generally followed guidance drafted by the European Commission in the development of their NASs and even non-member states look to the EU plan for guidance. While most NASs in Europe reference the Coordinated Plan, human rights are not considered as actively as, for example, economic benchmarks. This may be because European countries are assuming that their commitment to the EU documents themselves is sufficient representation of their human rights commitments.
- The three **Middle Eastern** strategies developed (from the United Arab Emirates (UAE), Saudi Arabia and Qatar) share goals and guidelines, providing some limited evidence for regional trends. They focus chiefly on data governance and highlight the importance of pro-business regulation regarding data sharing and portability as a central goal. Central to the strategies is making the country attractive for business and leaders in the development of AI technology. At the regional level, the League of Arab States has established an Arab Working Group on AI, chaired by Egypt which, among other things, is considering "developing an Arab AI strategy". Outside of these countries, Israel is perhaps the most advanced in terms of actions and measures taken in relation to AI governance, although much of the effort is driven toward research, with the establishment of a cross-governmental team in 2020 to devise recommendations for a policy plan to promote AI research and innovation activities in the country.
- Many of the strategies in the **Indo-Pacific region** focus on "human-centered" approaches to AI, outlining the concept of "making AI work for humans" as a key goal. Most of the strategies nod to potential ethical challenges raised by the development of AI

but see the development of ethical guidelines and inquiry as outside of the explicit scope of the overall strategy, instead establishing a separate working group or strategy for this focus. Some East Asian strategies highlight the need to articulate the purpose of AI beyond simply encouraging the development of AI as a central concern. Singapore emphasizes focusing on "benefits to citizens and businesses, i.e. getting AI to serve human needs, rather than developing the technology for its own sake". Japan and South Korea similarly focus on a human-centered approach in their wide-spread development plans. Australia and India engage most deeply with ethical risks and regulatory challenges but establish the development and articulation of ethical principles as separate goals from the main overarching strategy.

- Mauritius is the only country in **Africa** to have adopted a comprehensive NAS. The document makes minimal reference to human rights, however. While Egypt announced that it had adopted an NAS in 2020, it is not a comprehensive strategy as such but a short vision and mission statement, alongside a list of strategy pillars, strategy enablers and five priority sectors. The document similarly does not make any reference to human rights. In the absence of strategies, a growing number of governments are considering AI through other policy documents and bodies including South Africa's Presidential Commission on the Fourth Industrial Revolution, Zambia's Smart Zambia e-Government Master Plan (2018 – 2030, Kenya's Distributed Ledgers Technology and Artificial Intelligence Taskforce and Nigeria's National Agency for Research in Robotics and Artificial Intelligence. At the regional level, the African Union established a working group on Artificial Intelligence at the end of 2019 in order to study "the creation of a common African stance on Artificial Intelligence", "the development of an Africa wide capacity building framework" and the "establishment of an AI think tank to assess and recommend projects to collaborate on in line with Agenda 2063 and the UN Sustainable Development Goals".

Methodology

This report was commissioned by Global Affairs Canada with the aim of making an original contribution to current discussions regarding the application of the international human rights framework to the governance of AI. It seeks to catalogue and analyze current approaches to incorporating the international human rights framework into existing National AI Strategies (NASs), and to identify international good practices.

For the purposes of this report, the authors considered an NAS as any "set of coordinated government policies that have a clear objective of maximizing the potential benefits and minimizing the potential costs of AI for the economy and society". Strategies needed to have been officially adopted by a government (or group of governments) but did not need to be funded to be included in this definition.

The international human rights framework was understood to refer to all civil, political, economic, social and cultural rights set out in the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights.

The research for this report was undertaken through three primary sources of information:

- An analysis of all existing NASs published in the English language as of 1 March 2021 (and listed at Annex 2 to this report). A very small number of NASs published as of this date were not fully available in English, however the authors, where possible, obtained English language translations of all or part of the NASs, which were also used for the analysis;[1]
- A literature review; and
- A series of in-depth interviews with seven human rights defenders and leading experts in the field of AI governance.

An initial draft of the report was circulated for feedback among a group of experts, listed in the acknowledgements, before finalization, taking the feedback into account.

[1] There were four NASs that the authors were unable to review because they were unable to obtain an authoritative English translation: those developed in Poland (Założenia do strategii AI w Polsce, 2018), Cyprus (Εθνική Στρατηγική Τεχνητής Νοημοσύνης, 2020), Latvia (Informatīvais ziņojums "Par mākslīgā intelekta risinājumu attīstību", 2020) and Indonesia (Strategi Nasional Kecerdasan Artifisial Indonesia, 2020).

1.

What is a National Artificial Intelligence Strategy?

1.1 Defining Artificial Intelligence

Before looking at NASs, we should first define "artificial intelligence" (AI). While AI is a commonly used term, it is less commonly defined, even in NASs, and the distinction between algorithmic decision making and AI is also left unclear. Algorithms are processes or sets of rules to be followed by a machine to make a calculation or decision, or to solve a problem. This algorithmic decision making by machines can be used for a range of purposes, many fairly innocuous - such as analyzing traffic patterns to help decide where new roads should be constructed. Other uses, particularly when relating to humans, have proved more controversial, such as to determine an individual's likelihood of committing a criminal offence, or their eligibility for social welfare support.

AI is a related but more advanced technology, where a machine is not simply applying pre-determined algorithms to datasets to generate an output but mirroring human intelligence more broadly. A significant branch of AI, machine learning involves a machine, through trial and error, refining algorithms *itself*, thereby "learning" how to perform a particular task or function. AI involves the analysis of very large quantities of data, which is why questions of data privacy and data bias are so central when we think about its human rights impacts. The particular function performed by AI could be anything from generating results in a search engine that are most useful for the user to accurately predicting whether a person has a particular illness or disease. At present, most AI can only undertake a single task or a small range of tasks. The terms "true AI" and "artificial general intelligence" refer to a speculative form of AI that, in the future, would be able to undertake all human cognitive functions.

Given the significant potential impacts that AI can have upon human rights, human rights defenders often define the term broadly. David Kaye, for example, the UN Special Rapporteur on freedom of opinion and expression, has referred to AI as "shorthand for the increasing independence, speed and scale connected to automated, computational decision-making. It is not one thing only, but rather refers to a 'constellation' of processes and technologies enabling computers to complement or replace specific tasks otherwise performed by humans, such as making decisions and solving problems".[2]

Similarly, Mark Latonero has said that "it is useful to think of 'AI' as a catchphrase for a cluster of technologies embedded in social systems. This includes machine learning, natural language processing, computer vision, neural networks, deep learning, big data analytics, predictive models, algorithms, and robotics—all of which are intrinsically situated in the social contexts where they are developed and deployed".[3]

1.2 Defining National Artificial Intelligence Strategies

An NAS is a strategy, ordinarily developed by a government (although there are examples of governments adopting or endorsing NASs developed by other organizations) which sets out its broad approach to AI, specific areas of focus, and activities that it will undertake which relate to AI. In doing so, NASs attempt to coordinate government policies in order to maximize the potential benefits for the economy and society, while minimizing the potential costs. The first state to adopt an NAS was Canada, in 2017, and since then over 25 governments and regional intergovernmental organizations have published them, with many more in the process of development. The vast majority take the form of a published government document, with a smaller number of governments opting instead to launch dedicated websites or allocate certain amounts of government spending, while still calling their efforts a "strategy".

[2] UN General Assembly, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc. A/73/348, 29 August 2018, Para 3.
[3] Latonero, M., Governing Artificial Intelligence: Upholding Human Rights, Data & Society, 2018, p. 8.

Although this report is focused on NASs, it is important to mention briefly that governments are taking a range of different approaches as they make decisions on AI policy. Governments in a number of African states including Kenya, Ghana and Rwanda have been moving forward on trying to improve their capacity in the field of AI and grappling with the accompanying challenges. Some governments have commissioned reports on different aspects of AI or developed broader strategies on the digital economy or digital integration. The European Commission has published several reports, including an AI blueprint and a report on the safety implications of AI. In the UK, the government created the Centre for Data Ethics and Innovation to work on developing good governance approaches for emerging digital technologies. NASs represent only one component of how governments are approaching this question, but in states that have them they represent the comprehensive roadmap which other decisions are meant to follow, and so they are important in understanding a government's overall approach.

National and Intergovernmental NASs

Although the first NAS was adopted in 2017, policies relating to AI have been developed in states for many years before that. In 1998, for example, the Next Generation Internet Research Act was passed by the US Congress. However, the adoption of comprehensive cross-governmental strategies on AI has only occurred in recent years. The vast majority of NASs have been developed by high-income states, largely in North America, Europe and Oceania, with a smaller number from Latin America, Asia and Africa.

Outside of formal NASs, a number of regional and intergovernmental organizations have also sought to coordinate policy among their members. While this list is not exhaustive, at the regional level the European Union (EU) has attempted to coordinate policy on AI across EU member states. In its Coordinated Plan on Artificial Intelligence, the EU encouraged member states to produce NASs by mid-2019.[4] The EU has also produced its Declaration of Cooperation on Artificial Intelligence (2018),[5] Artificial Intelligence for Europe (2018),[6] Ethics Guidelines for Trustworthy AI (2019),[7] Policy and Investment Recommendations for Trustworthy Artificial Intelligence (2019),[8] and a White Paper on Artificial Intelligence (2020).[9] The Council of Europe launched an Ad hoc Committee on Artificial Intelligence (CAHAI) in 2019.[10]

Additionally, a number of Nordic and Baltic governments issued a Joint Declaration on "AI in the Nordic-Baltic region",[11] Canada and France published a statement calling for an international study group of artificial intelligence in 2018[12] (which was re-envisioned as the Global Partnership on Artificial Intelligence launched in 2020),[13] and the G7 has released the Charlevoix Common Vision for the Future of Artificial Intelligence.[14] The OECD also released

[4] European Commission, Coordinated Plan on Artificial Intelligence, December 2018.
[5] European Union, Declaration: Cooperation on Artificial Intelligence, 2018.
[6] European Commission, Artificial Intelligence for Europe, April 2018.
[7] Independent High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, April 2019.
[8] Independent High-Level Expert Group on Artificial Intelligence, Policy and Investment Recommendations for Trustworthy AI, June 2019.
[9] European Commission, White Paper On Artificial Intelligence - A European approach to excellence and trust, February 2020.
[10] Ad hoc Committee on Artificial Intelligence Terms of Reference, 2019.
[11] Nordic Council of Ministers, AI in the Nordic-Baltic region, May 2018.
[12] Canada-France Statement on Artificial Intelligence, June 2018.
[13] Joint Statement from founding members of the Global Partnership on Artificial Intelligence, 2020.
[14] G7, Charlevoix Common Vision for the Future of Artificial Intelligence, 2018.

its "Principles on Artificial Intelligence" in 2019[15], endorsed in the 2019 G20 AI Principles.[16] APEC has shown increasing interest in AI development, and the African Union ministers responsible for communication and information and communication technologies have established a working group on AI to study "the creation of a common African stance on Artificial Intelligence". The League of Arab States is "considering developing an Arab AI strategy".[17] The United Nations is also active in this space, with UNESCO developing a Recommendation for AI Ethics[18] and the UN High Level Panel for Digital Cooperation's roundtable 3C on AI.[19]

1.2.1 Purpose of National Artificial Intelligence Strategies

The reason that there has been a surge in the number of NASs in recent years is fairly straightforward: AI has become increasingly influential in different areas of life. From the public sector and public services to a range of private sector contexts, governments are competing to harness the economic and social benefits offered by AI. They are also keen to ensure that they remain ahead of the game and have a clear strategy in relation to the technology so as to maximize the potential benefits that AI brings and to minimize the potential costs. (What a government considers to be "benefit" or a "cost" will, of course, vary.)

Unlike many areas of government policymaking, however, AI is a cross-cutting issue with potential impacts in many different policy areas. As such, although there is often a lead department in charge, NASs invariably engage a range of different government departments, necessitating a cross-government approach to the development and implementation of the NAS.

By publishing a strategy, rather than simply keeping it as an internal point of reference, NASs also provide information to key stakeholders, often with different potential applications. Interviewees cited various uses of NASs reflecting the objectives of different stakeholder groups: civil society organizations, for example, noted that NASs are - or can be - useful accountability mechanisms, identifying government commitments and holding governments to account for them. Private sector interviewees, however, saw NASs as useful indicators for government policy which might affect them, such as regulation, opportunities for funding, and priorities for skills and talent development. A number of interviewees also noted that, whether the intention of governments or otherwise, the growing number of NASs meant that their contents were starting to contribute to international policy frameworks and norms to govern AI and that, in the absence of specific international instruments on the governance of AI, NASs helped to identify common values among governments.

1.2.2 Scope of National Artificial Intelligence Strategies

The development of NASs is a recent phenomenon, with governments taking different approaches; as such, it is unsurprising that the scope of these strategies varies and will likely continue to do so. The interviewees for this study noted the diversity of approaches when it came to NASs, not only in terms of their scope, but their length and level of detail, mirroring findings from analysis of the strategies themselves.

---

[15] OECD, Council Recommendation on Artificial Intelligence, May 2019.
[16] G20 Ministerial Statement on Trade and Digital Economy, June 2019.
[17] Ministry of Communications and Information Technology, Egypt Elected Chair of Arab AI Working Group, February 2021, available at: https://mcit.gov.eg/en/Media_Center/Latest_News/News/57187.
[18] UNESCO, First draft of the Recommendation on the Ethics of Artificial Intelligence, September 2019.
[19] Report of the Secretary General Roadmap for Digital Cooperation, June 2020.

A number of existing studies have sought to develop taxonomies of the different elements of existing NASs, most notably a taxonomy devised for a review undertaken for CIFAR (formerly the Canadian Institute for Advanced Research) by Tim Dutton (which he has utilized in his study for CIFAR[20] as well as other NAS-related publications,[21] a modified form of which was used by Anastassia Lauterbach)[22] and the World Economic Forum.[23] Together, these taxonomies suggest that there are between five and ten particular elements ordinarily covered in an NAS. Through the research for this study, however, an additional two elements found in a number of NASs were noted (one of which sets out the framing, vision or objectives of the NAS, while the other sets out how the strategy will be overseen or governed and the means by which other stakeholders will be involved in its oversight and implementation). In addition, one element found in existing taxonomies (data and digital infrastructure) was often, in practice, separated into two separate components (one dealing with data and one with infrastructure and its security). In total, therefore, this research found that there were up to thirteen elements that were commonly found in an NAS. These thirteen elements are not entirely discrete, and so some of these could be treated together (for example "ethics" and "regulation", or "talent" and "skills and the future of work").

The thirteen elements identified and used in this report are set out in the first column of Table 1, with the corresponding elements used in the other taxonomies with definitions across the remaining columns.

[20] Dutton, T., et al, Building an AI World: Report on National and Regional AI Strategies, CIFAR, 2018.
[21] See, for example, Dutton, T., AI Policy 101: An Introduction to the 10 Key Aspects of AI Policy, July 2018, available at: https://medium.com/politics-ai/ai-policy-101-what-you-need-to-know-about-ai-policy-163a2bd68d65.
[22] Lauterbach, A., Artificial intelligence and policy: quo vadis?, Digital Policy, Regulation and Governance, Vol. 21, No. 3, 2019.
[23] World Economic Forum, Centre for Fourth Industrial Revolution, A Framework for Developing a National Artificial Intelligence Strategy, 2019.

Table 1: Mapping elements of an NAS under different taxonomies

GPD-Stanford GDPI (this report)	CIFAR	Tim Dutton: AI Policy 101	Lauterbach	World Economic Forum
1. Framing, Vision and Objectives	n/a	n/a	n/a	n/a
2. Research	*Scientific Research*: The creation of new research centers, hubs, or programs in basic and applied AI research or a commitment to increase existing funding for public AI research.	*Basic and Applied Research*: To achieve new breakthroughs in AI theories, technologies, and applications, governments need to provide funding for basic and applied research. This includes both research grants and the creation of new research institutions. Example: the UK's Alan Turing Institute.	*R&D*: Governmental funding for R&D, including grants and creation of new research institutions, e.g. the UK's Alan Turing Institute.	*Key dimension 2*: Establishing a strong research environment and forging industry-academia integration.
3. Talent	*AI Talent Development*: Funding to attract, retain, and train domestic or international AI talent, including funding for chairs and fellowships or the creation of AI-specific Master and PhD programs.	*Talent Attraction, Development, and Retainment*: To conduct R&D in AI and deploy AI solutions in the public and private sectors, countries need a supply of skilled AI talent. Example: Canada's CIFAR Chairs in AI Program.	*Talent*: Ways to support talent acquisition to conduct R&D, e.g. Canada's CIFAR Chairs in AI Program.	n/a
4. Skills and the Future of Work	*Skills and the Future of Work*: Initiatives to help students and the overall labor force develop skills for the future of work, such as investments in STEM (science, technology, engineering, and mathematics) education, digital skills, or lifelong learning.	*Future of Work and Skills*: Advances in AI will both create and destroy jobs. To ensure that workers have the skills to compete in the digital economy, governments need to invest in STEM education, national retraining programs, and lifelong learning. Example: Denmark's Technology Pact.	*Future of Employment Skills*: Addressing the necessity to provide life-long education, e.g. Finland teaching 1% of the country's population the basic concepts at the root of artificial technology.	*Key dimension 3*: Preparing the workforce for the AI economy.
5. AI in the Government	*Industrialization of AI Technologies*: Programs to encourage private-sector adoption of AI technologies, including investments in	*Industrialization of AI Technologies*: AI has the potential to fundamentally transform multiple sectors and drive growth for decades	*Adoption of AI in Industries*: Governments investing in strategic sectors to boost AI ecosystems, e.g. China's investment in	*Key dimension 4:* Investing primarily in strategic sectors

	strategic sectors, funding for AI start-ups and small and medium-sized enterprises (SMEs), and strategies to create AI clusters or ecosystems.	to come. To encourage private sector uptake, governments are investing in strategic sectors and developing AI ecosystems and clusters. Example: Japan's Industrialization Roadmap.	self-driving cars and infrastructural regional clusters.	
6. AI in the Government	*AI in the Government*: Pilot programs that use AI to improve government efficiency, service delivery, and public administration.	*AI in the Government*: Likewise, governments are experimenting with ways to encourage the uptake of AI in the government. With the help of AI, it is possible to reform the public administration and make policy more effective. Example: UAE's Ministry of Artificial Intelligence.	n/a	n/a
7. Data	*Data and Digital Infrastructure*: Funding for open data partnerships, platforms, and datasets, as well as commitments to create test environments and regulatory sandboxes.	*Data and Digital Infrastructure*: Data is central to the ability of AI to work. As a result, governments are opening their datasets and developing platforms to encourage the secure exchange of private data. Example: France's Health Data Hub.	*Data*: Governments opening their data sets to encourage AI R&D and product development, e.g. France's Health Data Hub.	*Key dimension 1*: Providing a set of standardized data-protection laws and addressing ethical concerns.
8. Infrastructure and Cybersecurity	As above	As above	n/a	n/a
9. Ethics	*Ethical AI Standards*: The creation of a council, committee, or task force to create standards or regulations for the ethical use and development of AI. This area also includes specific funding for research or pilot programs to create explainable and transparent AI.	*Ethics*: Concerns over algorithmic bias, privacy, and security have raised a number of ethical debates. To mitigate harm, governments are looking to develop ethical codes and standards for the use and development of AI. Example: The EU's Draft AI Ethics Guidelines.	*AI Ethics*: Governments are trying to develop ethical standards for development of AI, e.g. The EU's Draft AI Ethics Guidelines.	*Key dimension 1:* Providing a set of standardized data-protection laws and addressing ethical concerns.

10. Regulation	n/a	*Regulations*: Every country is grappling with the question of whether (and how) to regulate AI. Currently, governments are focused on regulations for autonomous cars and autonomous weapons. Example: Germany's Ethics Commission on Automated and Connected Driving.	n/a	n/a
11. Inclusion	*Inclusion and Social Well-Being*: Ensuring that AI is used to promote social and inclusive growth and that the AI community is inclusive of diverse backgrounds and perspectives.	*Inclusion*: AI can both improve and worsen inclusion. Used properly, AI can bolster inclusion and help address complex societal problems such as poverty and hunger. Used improperly, AI can reinforce discrimination and disproportionately harm women and minorities. Example: India's #AIforAll Strategy.	*Inclusion*: Governments looking into how AI can address complex societal problems such as poverty, and/or harm or benefit women and minorities, e.g. India's #AIforAll Strategy.	n/a
12. Foreign Policy and International Cooperation	n/a	*Foreign Policy:* Geopolitics, development, and trade will all be affected by advances in AI technologies. To address ethical concerns and develop global standards, countries are beginning to consider mechanisms for the global governance of AI. Example: China's Global Governance of AI Plan.	n/a	*Key dimension 5*: Engaging in international collaboration.
13. Governance and Stakeholder Engagement	n/a	n/a	n/a	n/a

Some of these elements are more common than others, as has been noted in particular by CIFAR's first and second editions of its report, "Building an AI World: Report on National and Regional AI Strategies".[24] Almost all NASs - particularly those focusing on AI from an economic perspective - contain sections looking at research, talent and AI in the private sector. However, most NASs, particularly the more recent ones, contain other elements such as data and digital infrastructure, ethics, skills development, and international engagement.

Unsurprisingly, NASs tend to play on that particular state's strengths, such as existing talent and research, and the application of AI in that state's strongest sectors. Many seek to seize strategic opportunities for the development of AI, to build a competitive advantage, through building or recruiting skills and expertise on AI in the state, and supporting research institutions.

For the purposes of this report, it is in the section on "ethics" where human rights (or other values systems) are most often considered. While many NASs contain sections on ethics, fewer use human rights language explicitly (except frequent references to privacy), suggesting that ethical frameworks, rather than human rights frameworks, have found favor with governments when it comes to examining and responding to the potential impacts of AI on humans and society. As noted in section 4, there are, in fact, ways for international human rights law and standards to be considered when developing all elements of an NAS.

1.2.3 General areas for improvement of National Artificial Intelligence Strategies

While many NASs are comprehensive documents, there are a number of ways in which they can be improved and made more effective, in addition to improving the incorporation of international human rights standards. In particular, a common absence from NASs are clear goals and indicators of success, a criticism also made by Stiftung Neue Verantwortung in a review of NASs in 2018.[25] Connected to this, some interviewees highlighted a lack of specific policy commitments in NASs. Interviewees noted that NASs should be considered as mechanisms for accountability, and that the presence of indicators, specific commitments and deliverables were critical to enable civil society and other stakeholders to hold governments to account for their implementation.

Second, despite the fact that AI is a cross-cutting issue which has impacts on a broad range of areas of life, not all NASs are developed through a "whole of government" approach. Many NASs do not provide clarity over how different government departments were involved in their development or will be involved in their implementation, meaning that key considerations and expertise that resides in certain departments were not taken into account.

Third, many interviewees noted that there was a "future only" focus in NASs, with an exclusive focus on AI in the future and a lack of any landscaping or assessment of the current status of AI and its existing impacts. This absence was not only substantive but also procedural, with a common lack of clarity in the NAS - and even within government - on the existing structures, bodies and mandates that are relevant to the different areas of AI policy.

Fourth, some interviewees considered that NASs are often too focused on government exclusively and failed to set out how other stakeholders would be involved in the implementation of the NAS.

---

[24] Dutton, T., et al, "Building an AI World: Report on National and Regional AI Strategies", CIFAR, 2018; Kung, J., et al, "Building an AI World: Report on National and Regional AI Strategies: Second Edition", CIFAR, 2020.

[25] Heumann, Dr. S. and Zahn, N., "Benchmarking National AI Strategies: Why and how indicators and monitoring can support agile implementation", Stiftung Neue Verantwortung, 2018.

Finally, some interviewees also noted that AI as a global, data-driven technology is not something that any one government can address by itself but that many NASs nonetheless failed to take into account the need for international coordination and engagement. This criticism mirrors an argument made by, among others, Mark Esposito, who has highlighted how "data flows align with geographic boundaries only incidentally, not fundamentally" meaning that "[g]eopolitically, nation-states are sovereign entities; but in the digital economy, they are sovereign in name only, not necessarily in practice".[26] As such,

> "(...) [T]to frame the matter [AI] in strictly national terms is to ignore how AI is developed. Whether data sets are shared internationally could determine whether machine-learning algorithms develop country-specific biases. And whether certain kinds of chips are rendered as proprietary technology could determine the extent to which innovation can proceed at the global level. In light of these realities, there is reason to worry that a fragmentation of national strategies could hamper growth in the digital economy."[27]

[26] Esposito, M., "The case against national AI strategies", *Project Syndicate*, October 2018, available at: https://www.project-syndicate.org/commentary/case-against-national-ai-strategies-by-mark-esposito-et-al-2018-10.
[27] *Ibid*.

2.

The importance of human rights in National Artificial Intelligence Strategies

In the next sections, this report looks at whether and how NASs currently incorporate human rights and makes recommendations for how human rights should be incorporated into strategies in order to improve them and make them more rights-respecting. Before doing this, however, it is important to lay out in detail precisely why the incorporation of international human rights law and standards into NASs is so important.

States have an obligation under international human rights law to protect the human rights of all people within their territory and jurisdiction from violations, caused either by their own policies or practices, or the acts of third parties, such as businesses or other individuals. These obligations rest on a combination of international treaties and customary international law, and may be supplemented by regional instruments and domestic constitutional commitments. Human rights considerations are applicable to all areas of government policy and practice, including AI policy and AI applications used by governments and other entities in society.

AI has the potential to impact a wide range of human rights, both positively and negatively. As the UN Special Rapporteur on freedom of opinion and expression noted in his report to the UN General Assembly in 2018, "AI tools, like all technologies, must be designed, developed and deployed so as to be consistent with the obligations of States and the responsibilities of private actors under international human rights law".[28]

NASs represent a government's (or group of governments') comprehensive roadmap for how they intend to approach AI. This includes how it will be governed, how they will support innovation in the AI sector, and how they will address the impact on people's lives and on labor. If human rights are not integrated into this strategy, there is a significant risk that protection of rights will be uneven across various sectors. As implementation of strategies to improve AI training or to develop new technologies move forward, they may do so without fully addressing risks related to the violation of rights. For this reason, it is critical that part of the NAS is a strategy for ensuring that human rights are protected.

Yet, as we will see in the following sections, relatively few NASs deeply engage with the human rights impacts of AI. The most straightforward reason for this is that governments in some states simply do not place a high priority on the protection of human rights in their policymaking. China, Russia and the UAE, for example, have all developed NASs but all perform poorly on metrics measuring the degree to which human rights are protected in the state.[29] In states such as these, consideration of the international human rights framework is unlikely to be seen in many state policies and strategies, regardless of the policy area concerned.

Even among governments that do place a priority on the protection of human rights in policymaking, and despite their international human rights obligations, full consideration of human rights might stand in contrast to other goals and objectives that the government has when it comes to AI. Where a government's primary goal with respect to AI is to enhance economic competitiveness or to ensure geopolitical competition (particularly when it comes to the military applications of AI), consideration of human rights - through, for example, regulation or restrictions and limitations on certain uses of AI - may undermine the government's objective. As such, governments who have certain goals or objectives may decide to exclude or minimize the consideration of human rights in their NAS. In other cases, however, governments may have simply neglected to consider human rights implications of AI for the simple reason that they did not consider the human rights framework relevant to AI or were unsure of how to properly integrate these implications into their strategies.

---

[28] UN General Assembly, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/73/348, 29 August 2018, Para 19.
[29] See, for example, Freedom House's latest "Freedom in the World" report which gave China, Russia and the UAE scores respectively of 11, 20 and 17 out of 100.

It is important to note that states' obligations to protect human rights do not go away because they are unclear or inconvenient. There are situations where human rights may be legitimately restricted and this includes cases where there is a tradeoff between the protection of two rights or under other clearly defined conditions. This is even more reason to engage with human rights analysis early in the process of AI development. In the future, as hard decisions arise, it is important to have considered the implications of the technology and how the state will properly balance its obligations as they relate to human rights.

As an increasing number of proposals emerge to establish governance frameworks for AI, some have encouraged the development of new ethical frameworks for AI that are either designed to replace the human rights framework or in some cases to go beyond it. In a 2020 report on ethical and rights-based approaches to governance of AI, researchers from the Berkman Klein Center at Harvard University found eight key themes in existing AI principles: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and the promotion of human values.[30] While there is a high degree of overlap between these themes and human rights (particularly privacy and non-discrimination), the links between others and human rights is less clear (such as professional responsibility and human control). As such, some governments may consider that a human rights-based framework is insufficient and that a broader ethical framework that takes into consideration other issues is preferable. In other cases, governments may be advocating for ethical frameworks outside of the human rights framework, not because human rights do not go far enough but because they find them too restrictive.

It is important to note that nothing prohibits governments from going further than human rights frameworks do in protecting people's human rights. Furthermore, there are significant opportunities to incorporate newly emerging AI-specific ethical insights into human rights-based analyses of AI. There are persuasive reasons to use the existing human rights frameworks, rather than any new ethical principles alone, as a floor and framework for the regulation of AI. That is, human rights should be the foundation that any other governance approaches build on. As Eileen Donahoe and Megan MacDuffee Metzger have argued, the international human rights framework is well-suited to the task of ensuring that AI is developed, used and regulated for the benefit of individuals and societies for several key reasons.[31]

First, the international human rights framework puts the human person at the center of any assessment of AI and makes AI's impact on humans the focal point of governance. Second, the international human rights framework, through its broad spectrum of both substantive and procedural rights, covers the most pressing societal concerns about AI, such as non-discrimination and privacy. This makes it well-suited as the foundation for governance of AI.

Additionally, the international human rights framework establishes and clearly defines the roles and responsibilities of both governments and the private sector in protecting and respecting human rights and in remedying violations of them. As well as obligations on states via international human rights treaties, the UN Guiding Principles on Business and Human Rights sets out the role of the state and the responsibilities of the private sector when it comes to businesses' impacts on human rights. This is valuable in the context of a technology which will require oversight from both governments and private companies.

---

[30] Fjeld, J. et al, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI", Berkman Klein Center Research Publication No. 2020-1.
[31] Donahoe, E. and Metzger, M., "Artificial Intelligence and Human Rights." Journal of Democracy Vol. 30, 2019, pp. 115-126.

Finally, although interpreted and implemented in different ways around the world, the international human rights framework enjoys a level of geopolitical recognition and status under international law that any newly emergent ethical framework would be unlikely to match. Governments in states that do not comply with the framework risk criticism and even ostracization from the international community. This does not mean that all governments fully embrace these principles as guiding norms or apply them perfectly, but it is safe to say that human rights standards enjoy a high level of legitimacy and this is a crucial advantage.

All of this makes the existing human rights frameworks critical as governments move forward in the development of governance approaches to AI and as they consider how to responsibly innovate in this space. It is therefore critical that governments take seriously the question of protecting human rights as they outline their overall strategic vision for AI.

3.

Analysis of human rights in National Artificial Intelligence Strategies

To understand whether and how NASs currently incorporate human rights, we analyzed every strategy that had been formally adopted by a state before 1 March 2021 and which were either available in English or for which we were able to obtain translations.[32] This amounted to strategies from over 30 states and 2 regional strategies (one from the EU and one from the Nordic-Baltic states).[33] As a result of this analysis, we identified a number of patterns and themes relating to whether and how human rights had been incorporated into NASs.

3.1 Explicit references to human rights/the human rights framework

A majority of the strategies made explicit reference to human rights in their text. While this demonstrates a recognition among most governments that there is a human rights dimension to discussions around AI, the depth of subsequent engagement in the human rights framework and its application to AI varied considerably. Indeed, in most cases where human rights were referenced, this was a passing mention or simply a nod to the government's general commitment to human rights, with little or no further discussion of human rights in the strategy.

Some strategies did note the particular importance of human rights as a component of their strategy (e.g. Germany)[34] and a small number did go into some depth as to how human rights were part of the core of their approach to AI (e.g. Denmark and the Netherlands). In general, however, strategies did not deeply engage with human rights as a framework for structuring their strategies. This finding is in line with previous work on NASs, which has found a lack of leadership on human rights in existing strategies.[35] Even where there was deeper engagement on the human rights dimension of AI, this was often only (or largely) in sections dealing with ethics, with far less attention paid to the application of human rights in other parts of the strategy. Many strategies did not mention human rights at all as a framework for the government's approach to AI.

It is important to note, as this report mentions later under section 3.6, that the fact that an NAS references human rights is far from sufficient to ensure that the government's approach and actions will in fact ensure that human rights are respected in practice. Indeed, some of the states whose NAS reference human rights have come under significant criticism from the international community over human rights violations in AI-related policy areas, such as the surveillance and the use of personal data.

Our analysis also revealed notable regional variation, with the vast majority of countries explicitly mentioning human rights and the human rights framework being in Europe. While many East and Southeast Asian states had developed an NAS, these did not use the human rights framework, sometimes referencing instead to alternatives, such as ethical or "human-centric" approaches. This may, in part, reflect variation across regions when it comes to using human rights as a framework for policy making, as well as varying levels of commitment to the international human rights framework more broadly.

Among European states, while most did reference human rights explicitly in their strategies, a small number did not. In the case of Europe, however, this may not necessarily represent a lack of commitment among governments in these states to the human rights framework. Interviewees suggested that in some of these states, human rights are often assumed to form the

[32] See above, note 1.
[33] A full list of all NASs reviewed can be found in Annex 2.
[34] In the case of the Russian strategy, it should be noted that a limited scope of human rights was highlighted, primarily the right to work.
[35] Cussins Newman, Jessica, *Towards AI Security: Global Aspirations for a More Resilient Future*.

foundation of policy whether or not it is explicitly stated. In other words, some European states may not feel that they need to explicitly discuss human rights in their strategies, because it is assumed as a baseline. This may be especially true in EU member states since the EU's regional AI strategy, which has strong links to the NASs developed by EU member states, does explicitly embed itself in a human rights framework. Additionally, the EU has published Ethics Guidelines for Trustworthy Artificial Intelligence that include provisions meant to protect human rights, including specific provisions related to privacy, diversity and non-discrimination, and the right to redress. The Lithuanian NAS, for example, explicitly mentioned these guidelines as forming the foundation for any ethical principles for EU member states going forward. That said, and as this report looks at later, if EU member states are excluding deep engagement with human rights because they are assumed as the baseline, this is still problematic.

Examples of Explicit References to Human Rights / the Human Rights Framework

"As a diverse, innovative nation, we will decide what impact this technology will have on human rights, on people's lives and on our democratic values." (Luxembourg)

"The government's basic task (which is to guarantee the fundamental and human rights of all people) will be intensified markedly as authorities can now respond to people's needs digitally, independent of time and location." (Finland)

"The basic principles of the development and use of artificial intelligence technologies, the observance of which is obligatory during the implementation of this Strategy, include (…) the protection of human rights and liberties: ensuring the protection of the human rights and liberties guaranteed by Russian and international laws, including the right to work, and affording individuals the opportunity to obtain the knowledge and acquire the skills needed in order to successfully adapt to the conditions of a digital economy." (Russia)

"Artificial intelligence that is developed and used in Norway should be built on ethical principles and respect human rights and democracy." (Norway)

3.2 Ethical or human-centric approaches as alternatives to the human rights framework

As noted above, our analysis found that, while some strategies did not mention human rights specifically, they nonetheless referenced ethical frameworks, ethical approaches, or human-centered approaches to AI. Often, however, there was no clear indication of the inspiration of these frameworks or approaches, nor how they would be developed, implemented or applied.

That being said, where greater detail was provided on these frameworks and approaches, certain human rights concepts and principles – such as privacy – were sometimes mentioned without referring to them being human rights. This implies that certain human rights are to some extent being considered, albeit not through the traditional human rights framework.

The debate around whether a purely human rights framework or a broader (or different) ethical framework should be used when considering the governance of AI goes beyond NASs. There is a wider debate in AI governance circles among governments, the private sector and other relevant stakeholders on this question. It is fair to say that there are indeed arguments against using a purely human rights-based framework when it comes to AI on the basis that it is too limited and fails to fully consider all the different impacts that AI can have on humans and societies.

Human rights-based frameworks and ethical frameworks are not, however, necessarily wholly different. As noted above in section 2, a 2020 report on ethical and rights-based approaches to governance of AI, researchers from the Berkman Klein Center at Harvard University found eight key themes in existing AI principles that have been developed: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and the promotion of human values.[36] This research showed that there is degree of overlap between alternative frameworks for governing AI and the human rights framework (particularly when it comes to privacy and non-discrimination). Certain aspects of these alternative frameworks, however, have fewer clear links to the human rights framework (such as professional responsibility and human control). As such, it may be that governments in some states considered that a human rights-based framework was insufficient, and that a broader ethical framework that takes into consideration other issues was preferable. As discussed in section 2 of this report, however, there are strong arguments in favor of using the human rights framework as a base from which to build out any framework for AI governance.

In addition to "ethics", another term that appeared in many NASs is "human-centric" AI (see, e.g., Singapore). Where the term was used, however, it was rarely defined clearly. The concept of "human-centric" hints at some aspects of a human rights-respecting approach, but it is hard to evaluate without clear definitions or descriptions. The term could mean an approach which is fully focused on the rights of humans as a priority over all else, or it could simply mean an approach that maintains human involvement at all levels of AI processes, or it could mean simply that humans should be the focus of policy but without clear attention to their rights. Each of these represents a very different commitment. In the absence of greater clarity among those NASs that use the term, it is not possible to determine to what extent the reference demonstrates a commitment to human rights. As with ethical frameworks and approaches, therefore, there are strong arguments - set out in section 2 of this report - as to why an explicit human rights framework should be preferred.

Examples of Ethical or Human-Centric Approaches as Alternatives to the Human Rights Framework

"Establish a charter of ethics for Intelligent IT to minimize any potential abuse or misuse of advanced technology by presenting a clear ethical guide for developers and users alike." (South Korea)

"Singapore will also apply multidisciplinary and human-centered approaches to study the systemic risks and long-term impact of AI, and develop potential solutions to address them. Risk assessment in AI development should not be narrowly confined to the engineering disciplines, but also include sociologists, ethicists, economists, lawyers and policy makers. Today, Singapore's universities are actively studying the societal implications of AI, and we will tap on their expertise." (Singapore)

"The development of the required ethics framework will involve specifying what is meant by the term human-centred, identifying ethical AI objectives and development directions and integrating these in a reliable and robust AI framework, as well as developing techniques for asserting and enforcing the rules of ethics." (Hungary)

[36] Fjeld, J. et al, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI", Berkman Klein Center Research Publication No. 2020-1.

3.3 Prioritization of specific human rights

Beyond explicit mentions of human rights as a framework, a number of strategies did mention and engage with specific human rights and the impact of AI on these rights. There were some particular trends in how certain human rights were prioritized or given particular focus and attention. Of the strategies which engaged with specific human rights, the right to privacy was the most commonly mentioned. As noted above, some strategies did also reference privacy as a principle but without framing it explicitly as a human right or part of the human rights framework. Because of this, and the fact that privacy as a general concept varies across different regions and contexts, references to privacy alone in an NAS, however, could not be considered as comparable to a broader human rights approach. Another right that is mentioned quite often is the right to equality or to non-discrimination. This is often engaged less deeply than privacy but nonetheless appears in a range of NASs.

Strategies vary on the types of specific human rights that they mention or emphasize. Some – for example, the Russian strategy – focus most of their ethical concern on issues like the right to work or the right to education and how AI will impact this. Others, like the French strategy, are focused more on problems of transparency, accountability and anti-discrimination and how these may be impacted by or amplified by AI. Still others, like the Indian strategy, focus the bulk of their ethical concerns on ensuring inclusion and that the benefits of AI are distributed across society. This reflects, to some extent, the particular concerns of the states themselves, as well as their values. That said, all of these concerns will likely pose challenges in all states and a focus on only a small subset of the potential impacts of AI leaves governments open to problems down the road. For example, governments which choose to focus on the impact on the right to work but without a clear focus on challenges related to equality or lack of discrimination may find themselves turning to AI solutions which, on average, address their concerns about labor but entrench existing disparities in other ways or even create new ones.

Table 2: Mapping human rights referenced in NASs

Human Rights Mentioned	States/Regional Organizations[37]
The right to privacy	Australia, Belgium, China, Czech Republic, Germany, India, Italy, Luxembourg, Malta, the Netherlands, Norway, Portugal, Qatar, South Korea, Spain
The right to equality / non-discrimination	Australia, Belgium, Czech Republic, Denmark, Estonia, EU, France, Germany, Italy, Malta, the Netherlands, Norway, Serbia, Spain
The right to an effective remedy	Australia (responsibility and ability to hold humans responsible), Denmark, Malta, the Netherlands
The rights to freedom of thought, expression and access to information	France, the Netherlands, Russia
The right to work	France, Russia

[37] In all cases the NAS itself was evaluated, with the exception of Australia and Malta which each has a dedicated AI ethics document which their NAS points to for guidance on these issues, and which was used in evaluation for these two states.

3.4 Discussion of human rights issues without considering risks

Our review noted many instances whereby a strategy, while not mentioning a specific human right, nonetheless engaged with an issue that posed risks to human rights. For example, a number of NASs included substantial analysis, often an entire section, on the implications of AI on the future of work. In some cases, strategies even included fairly specific plans of action around education, retraining and other factors relevant to protecting the right to work, yet almost invariably did so without making any reference to the right to work specifically and engaging with the risks to that right posed by AI.

To provide another example, some strategies referenced the need for algorithmic transparency and fairness but did so without applying the human rights framework, therefore failing properly to engage with risks to human rights such as the right to an effective remedy or the rights to equality and non-discrimination.

Risks to human rights, where these were discussed, were often analyzed superficially and without reference to the human rights framework as a means for identifying those risks. In some instances, strategies did mention risks, only to immediately minimize them and move on to focus entirely on positive benefits in the remainder of the strategy. For example, the initial Finnish NAS starts by noting that "[t]he bleakest predictions indicate that the amount of jobs and work available to people will decrease, the meaningful content of work will decline and the labor market status and earnings trend of employees will become more uncertain". However, rather than engage here (or elsewhere) with risks to the right to work (or to just and favorable conditions of work), it simply proceeds in the next sentence to suggest that this is unlikely in reality with no evidence presented in support.

This is not to say that NASs should present only worst-case scenarios when it comes to the potential impacts of AI, nor that there should be no acknowledgement of the potential benefits. However, it is critical that strategies engage deeply and meaningfully with the risks in order to ensure that they in fact reap the potential benefits. Indeed, in follow-up reports to the initial Finnish NAS, there was much more explicit engagement with challenges, particularly as related to trust, security and the nature of work.

3.5 Lack of specificity or incentives

As illustrated by the example of "human-centric" approaches above, a common finding revealed by our review was a lack of depth and specificity on how human rights (or even ethics) should be protected. While almost all strategies highlighted the need to ensure that potential harms were mitigated against (either by ensuring the protection of human rights or by using an ethical approach), strategies largely failed to set out any specific details of *how* this should be done in practice. This absence of detail stood in stark contrast to other parts of these strategies which were often quite specific and detailed (for example, about the types of investments that will be made in the areas of research and how talent would be developed).

There are some notable exceptions to this general finding. In particular, the EU Communication on Artificial Intelligence, in its section on ethics, makes specific commitments to the development of ethical guidelines grounded in human rights, as well as further research and commitments to issues such as the explainability of AI and the need for effective data protection.

The Danish strategy also stands out for both the specificity of its commitments around human rights and its integration of these rights throughout the strategy itself. The Danish NAS outlines

ethical principles for AI which are grounded in human rights principles: self-determination (making people's autonomy paramount in the use of AI), dignity (preventing AI from harming people or disrupting human rights or democratic processes), responsibility (responsibility for mistakes should be able to be placed onto human beings), explainability (AI decisions should be able to be explained and supported), equality and justice (preserving diversity and preventing bias in algorithms), and development (development of AI should be ethically responsible and oriented towards maximizing societal progress). It then makes clear commitments as to how these principles would translate into practical actions, including establishing a data ethics council, improving the clarity of legal responsibility around AI, and improving the ethical use of data in business. The Danish NAS also aims to make Denmark a leader on these issues in order to try to ensure a human rights-based approach elsewhere through international cooperation.

Even within a strategy, though, there was sometimes inconsistency. For example, India's strategy lays out very specific commitments to tackling risks to privacy but makes no such commitments to other potential issues despite acknowledging risks such as algorithmic bias.

3.6 Words vs practice in National AI Strategies

The Danish strategy also stands out for both the specificity of its commitments around human rights and its integration of these rights throughout the strategy itself. The Danish NAS outlines ethical principles for AI which are grounded in human rights principles: self-determination (making people's autonomy paramount in the use of AI), dignity (preventing AI from harming people or disrupting human rights or democratic processes), responsibility (responsibility for mistakes should be able to be placed onto human beings), explainability (AI decisions should be able to be explained and supported), equality and justice (preserving diversity and preventing bias in algorithms), and development (development of AI should be ethically responsible and oriented towards maximizing societal progress). It then makes clear commitments as to how these principles would translate into practical actions, including establishing a data ethics council, improving the clarity of legal responsibility around AI, and improving the ethical use of data in business. The Danish NAS also aims to make Denmark a leader on these issues in order to try to ensure a human rights-based approach elsewhere through international cooperation.

Even within a strategy, though, there was sometimes inconsistency. For example, India's strategy lays out very specific commitments to tackling risks to privacy but makes no such commitments to other potential issues despite acknowledging risks such as algorithmic bias.

4.

Recommendations for incorporating human rights into National Artificial Intelligence Strategies

As discussed in section 1, the goal of NASs are generally to maximize the benefits of AI technology for a state while minimizing the risks. This has meant that much of the focus has been on economic impacts and how to harness AI to benefit a state's economy and competitiveness. There has also been insufficient focus on asking whether AI is an appropriate solution for a given case. This innovation and economic benefits-centered approach is consistent across most NASs. What is less consistent is how they deal with the potential risks to human rights and challenges posed by AI. As we have discussed above, a small number of strategies do deeply engage with these questions but these are the exception rather than the rule.

Despite the variance in how existing NASs address human rights, states have obligations under international human rights law to respect, protect and promote them. These obligations are often also reinforced and complemented by regional and national human rights frameworks. It is therefore incumbent upon governments to ensure that they fully consider the human rights dimensions of all areas of public policy. This is no less true when it comes to the issue of AI.

The approach that different stakeholders – particularly governments and the private sector – take towards AI will greatly influence whether the development and use of AI in society will bring benefits or harms to human rights. An approach which focuses solely on the economic aspects of AI, for example, may fail to consider broader societal and individual impacts, leaving human rights, including economic rights, at risk in the long term. An approach which integrates human rights, on the other hand, will be far more likely to identify and mitigate risks. One critical place to ensure that a comprehensive, consistent and human rights-respecting approach is taken is during the development of the NASs.

Based on the current state of NASs globally, there are a number of suggestions for steps governments can take to ensure that their NAS lays the groundwork for a human rights-respecting approach to AI policy. These are outlined in more detail in the assessment tool which is associated with this report (Annex 1), but they fall into a few core categories.

4.1 Include human rights explicitly and throughout the Strategy

Thinking about the impact of AI on human rights and how to mitigate the risks associated with those impacts should be core to an NAS. Protection of human rights should be laid out explicitly as a foundation and the strategy should return to engage with specific rights throughout all segments. Simply mentioning that human rights must be protected or including them in a separate ethical section does not set the state up well to consider respect for the protection of human rights as critical at all stages. It is also not sufficient to engage only with the concerns which seem most pressing to the government at that moment. All potential risks are likely to touch all states, and acknowledgement of some risks at the expense of others leaves governments open to problems later down the road.

Even for governments which may consider human rights to be assumed as a baseline, it is important to be explicit and specific about the commitment to the protection of human rights in this context within the strategy. This is important both as a signal to other governments which may look to existing strategies in developing their own, as well as for cementing human rights as a core consideration along with other considerations such as innovation, economic success and competitiveness.

Throughout a strategy, each section should consider the risks and opportunities AI provides in relation to human rights, as well as include a specific focus on at-risk, vulnerable and marginalized communities. For example, a section on the impact of AI on the healthcare system should discuss both the potential positive benefits of AI improving healthcare access or health

outcomes (and positively impacting the right to health), while also acknowledging the risks of privacy and potential discrimination that can come if this is implemented without due consideration for the quality of data or an effective data protection framework.

As noted in section 2, there are thirteen broad elements that are commonly found in NASs. As we outline in more detail in the context of the assessment tool (Annex 1), there are ways to integrate human rights considerations into all of these elements. It is important not to confine the discussion of human rights impacts to a separate ethics section. If an NAS is viewed as a blueprint for how a government is going to develop policies around AI, it is important to integrate a discussion of the potential risks and opportunities for human rights throughout all components of that strategy to ensure that as policies are enacted, research is funded, etc. these considerations are already front and center.

Good Practice Examples

"Norwegian society is characterised by trust and respect for fundamental values such as human rights and privacy. The Government wants Norway to lead the way in developing and using AI with respect for individual rights and freedoms." (Norway)

"For this reason, this Strategy will bring citizens closer to the debate on the most relevant aspects of AI and its governance, promoting the creation of national and international forums for dialogue. These spaces will address fundamental issues, such as ethics in AI systems and the impact of this technology on Human Rights and public freedoms." (Spain)

4.2 Outline specific steps to be taken to ensure human rights are protected

As strategies engage with human rights, they should include specific goals, commitments or actions to ensure that human rights are protected. For example, rather than simply saying that the government will develop ethical guidelines for the use of AI, a government should commit to a process through which they will do this and emphasize that any guidelines must have human rights as their foundation. The strategy could set out who will help develop these guidelines, specify that they will be based on human rights, set a timeframe for the circulation of guidelines for consultation, and clarify the status that such guidelines will have. These details could be contained within the strategy itself or it could be accompanied by an action plan which sets out these details. The second approach allows a strategy to be a long-term, higher-level document but with regularly updated and published action plans, both on plans for implementation as well as reviews on progress.

Making specific commitments rather than general ones can help to hold governments accountable for the commitments they are making. If there are clear benchmarks, it becomes possible for people to evaluate how well these are being achieved. Many of the strategies to date have very specific commitments when it comes to research or to deploying AI in certain fields. This same specificity and detail should be used when dealing with the impact of AI on human rights.

In setting out the steps to be taken, governments should always bear in mind the need to take into account the different impacts that AI has upon different groups, particularly those vulnerable to discrimination - such as women, persons with disabilities, and racial and ethnic minorities - as well as the intersectionality between different potential forms of vulnerability. Different goals, commitments or actions may be needed to ensure that risks to human rights

faced by particular groups are sufficiently mitigated. As noted below, members of these groups and experts in the particular risks faced by these groups should also be included in the process of developing NASs.

Good Practice Example

"Actions:

- Commissioned by the government, the Scientific Council for Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid, WRR) will investigate the impact of AI on public values.
- Commissioned by the Research and Documentation Centre (Wetenschappelijk Onderzoek- en Documentatiecentrum) (part of the Ministry of Justice and Security), Utrecht University has started a research programme under the name 'Legal aspects of decision-making algorithms' (Juridische aspecten van algoritmen die besluiten nemen). This exploratory study involves a study of five cases to identify the key opportunities and risks of decision-making algorithms over the next five to ten years and how they relate to the existing legal frameworks (and the values that lie behind them). The cases to be examined are: the self-driving car, P2P energy markets, judges, 'doenvermogen' (self-efficacy) and content moderation on platforms.
- On behalf of the Minister for Legal Protection, Tilburg University is conducting research into the risks to our privacy associated with the use of facial recognition technology and into possible measures to limit these risks. This study is expected to be completed by the end of 2019." (The Netherlands)

4.3 Build in incentives or specific requirements to ensure rights-respecting practice

Governments should take steps within their strategies to incentivize human rights-respecting practices and actions across all sectors, as well as to ensure that their goals as regards the protection of human rights are fulfilled. This could include tying research funding to the outcome of human rights impact assessments or requiring courses on human rights and AI in training programs. It might also include requirements when it comes to government procurement of AI or regulation of the use of AI in the public and private sectors. As with attempting to be specific, the goal here is to help ensure the translation of good principles on paper into good practices in the real world. Building in incentives or requirements can help ensure that the values don't remain on the page.

This should also ideally include creating processes or institutions which will help to ensure the protection of human rights even as new challenges and concerns related to AI emerge. Because of the pace at which technology is changing, it is likely that some of the specific challenges AI poses are not yet known. In order to prepare to address these in the future, governments should ensure that their strategies tackle not only known problems but ensure that there are structures for addressing problems that can adapt over time. Some of these institutions and structures could be ones which already exist, such as national human rights institutions, data protection authorities, or equality bodies. In these cases, further capacity-building to ensure that they are able to exercise their existing mandates while taking into consideration the impacts of AI may

be sufficient. Where there are still gaps, the creation of new bodies may be necessary, such as the United Kingdom's Center for Data Ethics and Innovation.[38]

Exactly what this looks like in practice may vary substantially by state. There are a range of ways to incentivize the protection of human rights including everything from direct regulation to official guidance and capacity-building for AI developers to financial incentives. There is no one perfect solution for every state but each government should take appropriate measures to ensure that the strategies goals' relating to human rights protections are fully realized.

4.4 Set out grievance and remediation processes for human rights violations

However deep the analysis of potential human rights impacts stemming from AI and however comprehensive the actions put in place to mitigate risks to human rights, violations and adverse impacts upon human rights are nonetheless a distinct likelihood. Recognizing this, the international human rights framework includes a right to an effective remedy when there has been a human rights violation: the UN Guiding Principles on Business and Human Rights are clear that governments have an obligation (above that of companies' own responsibilities) to ensure that victims of human rights violations caused by the actions of business enterprises are able to access effective grievance and remedial processes.

An NAS should look at the existing grievance and remedial processes available for victims of human rights violations relating to AI and determine whether they are sufficient. These might include general pieces of legislation protecting human rights which give victims the ability to bring grievances through court processes or through other human rights mechanisms such as national human rights institutions. They might include specific processes, such as a data protection authority which can bring proceedings against actors which misuse personal data, or courts or tribunals that can hear cases of discrimination (including discrimination caused by AI). However it may be that existing processes (including their legislative underpinnings) need revision in light of the particular nature of AI as a technology or capacity-building of those involved in these processes so that they are able to receive complaints which involve AI. There may also be a need to ensure that individuals who are victims of AI-related human rights violations are aware of the mechanisms by which they can pursue grievances, that the potential remedies available are effective in addressing the harms that are caused, and that private enterprises who develop or use AI are appropriately incentivized to provide their own grievance and remedial processes.

4.5 Recognize the regional and international dimensions to AI policy

As a global technology, no one government is likely to be able fully to manage the risks to human rights posed by AI alone. The companies developing the technology used in that state may be based in other states and therefore difficult to regulate; there may be partnerships among different governments, research institutions or other actors on AI across national borders; and governments may want to align their own AI policies to those of other states to enhance their attractiveness and competitiveness. For these and other reasons, governments should be looking at regional and global forums and processes which are also setting common policies, guidelines or standards relating to AI. These include the work being undertaken by the Council of Europe and UNESCO.[39]

---

[38] The Centre for Data Ethics and Innovation is an independent advisory body, set up by the UK government with a mandate to investigate and provide advice on how the UK could maximise the benefits of data-driven technologies.

[39] In September 2019, the Council of Europe established an Ad Hoc Committee on Artificial Intelligence whose remit includes examining the feasibility of a legal framework for the development, design and application of AI. In March 2020, UNESCO appointed an international expert group to draft global recommendations on the ethics of AI.

From a human rights perspective, it is critical that governments looking at international cooperation on AI do not simply set out how they intend to engage in these forums and processes but proactively promote approaches and outcomes at them which are consistent with the human rights framework. Failure to do so risks these regional and global approaches and outcomes failing to appropriately protect human rights, and given that these approaches and outcomes will likely influence the development and use of AI in other states, risks undermining any efforts made at the national level. NASs should therefore clearly identify relevant regional and global forums and processes relating to AI, and the means by which the government will promote human rights-respecting approaches and outcomes at them through proactive engagement.

Good Practice Example

"We will expand international, bilateral and multilateral cooperation on AI, for example within the G7 and the G20. International cooperation has long been a key feature of cutting-edge research projects, and most scientific communities – including the AI community – are already very well connected across borders, as they share common research interests. This cooperation and these networks are to serve as the foundation for European research facilities to engage in further cooperation projects and develop outstanding solutions that will be successful internationally. We need to work with the nations leading this field – for example the US, Canada, Israel and some Asian countries – to conduct joint bilateral and/or multilateral R&D activities on the development and use of AI. This also includes cooperation between companies from different countries which are part of the same global value chains. German diplomatic missions and the German Houses of Research and Innovation can be used for this type of cooperation. Germany Trade & Invest GmbH (GTAI) will actively disseminate the importance attached to AI in Germany as part of its foreign trade and investment efforts. This will help improve the visibility of German providers of AI solutions and make foreign companies aware of the opportunities for investment and cooperation that exist in Germany.

When it comes to developing common guidelines, we advocate taking a multilateral approach by using existing forums such as the OECD, G7, G20 and the United Nations." (Germany)

4.6 Include human rights experts and other stakeholders in the drafting of NAS

Finally, when drafting an NAS, the government should ensure that experts on human rights and the impact of AI on human rights are a core part of the drafting process. Even if the focus of the document is on innovation, including human rights experts will ensure that key risks or opportunities relating to AI and human rights are not missed. Just as the government would bring in experts on economics or the labor market to help with drafting to ensure that good strategies are developed around those topics, they should bring in experts on human rights to ensure that they are well-integrated into the overall strategy. This should include not only general human rights organizations, but also a broad range of civil society organizations and other stakeholders representing communities that may be adversely affected by AI or benefit particularly from certain applications and with expertise in digital technologies (including groups such as women, persons with disabilities, and racial and ethnic minorities).

5.

Regional developments and trends in Artificial Intelligence governance

While sections 3 and 4 of this report look at the state of play when it comes to National Artificial Intelligence Strategies (NASs) globally, as the number of published NASs increases, trends at the regional and sub-regional level are emerging. These trends relate not only to the content of those NASs but also their framing and the extent to which human rights are (or are not) considered.

At the same time, while NASs provide an opportunity for a government to set out a holistic approach towards AI, including its governance, they are not the only means through which a government can steer the direction of AI. Developing a comprehensive strategy is a process which is both time and resource intensive and some governments, particularly those with limited resources or in states where the use of AI is at a more nascent stage, may feel that it is not appropriate to develop a strategy. Many of those governments have nonetheless taken actions or measures in relation to AI which sufficiently influence the development, use and governance of AI such that human rights can - and should - be considered in a way comparable to that outlined in the earlier sections of this report.

In this section, we take a look at some of the regional trends from a human rights perspective that are beginning to emerge when it comes to AI governance, both in regions where many NASs have been adopted and those where alternative approaches are more common.

5.1 North America and the Caribbean

Within North America, both Canada and the United States have adopted NASs.[40] In Mexico, a report, "Towards an AI Strategy in Mexico" was published in 2018 by Oxford Insights C Minds. The report was funded by the British Embassy in Mexico although the Mexican government was also involved in its development. This was followed, in 2020, by the publication of the Mexican National Agenda for Artificial Intelligence, developed by the IA2030Mx Coalition, a multistakeholder coalition of professionals, academic institutions, companies, startups, public agencies and other actors in Mexico's digital ecosystem. Given its status as a document not published by the government of Mexico, it contains no commitments but instead recommendations to the government. The report does, however, take an explicit human rights-based approach to identifying risks and challenges posed by AI, with impacts upon freedom of expression, equality and non-discrimination, and privacy extensively discussed.

In the absence of a formal governmental strategy, the government of Mexico has nonetheless taken action to steer the development and use of AI within government, publishing "Principles and Guidance for Impact Analyses for the Development and Use of Systems Based on Artificial Intelligence In the Federal Public Administration" in 2018.[41] The principles and guidance explicitly note that they intend, among other things, to ensure respect for human rights. The principles themselves include a number of human rights considerations and the guiding questions include a number which examine potential impacts on the rights to privacy and non-discrimination.

---

[40] It should be noted that while the Trump administration did refer to its American AI Initiative as a National AI Strategy, the US Congress subsequently passed a resolution calling for the development of a National AI Strategy, suggesting some tension in the understanding of the strategy between Congress and the former administration. Although not a NAS, the US National Security Commission has also published a report on artificial intelligence which contains many recommendations relevant to National AI Strategies: National Security Commission on Artificial Intelligence: Final Report, 2021, available at: https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

[41] Government of Mexico, Principios y Guía de Análisis de Impacto Para el Desarrollo y Uso de Sistemas Basados en Inteligencia Artificial en la Administración Pública Federal, 2018, available at: https://www.gob.mx/innovamx/articulos/guia-de-analisis-de-impacto-para-el-desarrollo-y-uso-de-sistemas-basadas-en-inteligencia-artificial-en-la-apf.

As of 2021, no Caribbean country has adopted a NAS, nor have there been any commitments by governments in the region to develop one. Instead, it is possible that Caribbean governments will develop a regional approach to AI. The most significant initiative in this regard is the Caribbean Artificial Intelligence Initiative, established in 2020 under the auspices of the UNESCO Cluster Office for the Caribbean with the objective of developing "a sub-regional strategy on the ethical use of AI.[42] As of the time of writing of this report, however, no further detail of the development of this strategy has been made publicly available, nor the extent to which human rights will be considered.

5.2 Central and South America

Of the Central and South American countries, only three have published NASs: Argentina, Colombia and Uruguay, making the identification of regional trends difficult. That being said, all three explicitly emphasized the importance of ensuring respect for human rights throughout the text. A number of other governments have committed to developing a strategy, including Brazil and Chile. Brazil has already developed a Digital Transformation Strategy and a National Plan on IoT; neither of these references human rights.

There are also initiatives taking place in other parts of the region. Costa Rica, for example, is considering AI as part of broader governmental strategy. In 2018, a High-Level Commission for Digital Government of the Bicentennial was established to act as an advisory body for the development of a national strategy aimed at the implementation of "the digital government public policy". At the same time, the government also published its own Digital Transformation Strategy towards Costa Rica's Bicentennial 4.0 and, in 2020, a National Code of Digital Technologies. None of these documents, however, references human rights considerations.

5.3 Europe

A significant number of European countries (24) have adopted NASs and many others are in the process of developing one, in part because of the EU's push in its Coordinated Plan on Artificial Intelligence. European countries have generally followed guidance drafted by the European Commission in the development of their NASs. The Commission's Coordinated Plan on Artificial Intelligence,[43] Ethics Guidelines for Trustworthy AI,[44] Declaration of cooperation on AI,[45] and other initiatives are consistently cited in European strategies as framing guidance.

For EU member states, publishing their requisite strategies in line with EU guidance is a requirement of the Coordinated Plan on AI; however, non-member states are also looking to the EU plan for guidance. Serbia, for example, cites its position in the EU accession process as a key factor in determining the goals of their strategy: "[the strategy] seeks to provide the necessary extent of compliance with the European Union, which will enable full integration into the European Research Area and closer cooperation."

---

[42] UNESCO, UNESCO Caribbean Artificial Intelligence Initiative, available at: https://en.unesco.org/caribbean-artificial-intelligence-initiative.
[43] European Commission, Coordinated Plan on Artificial Intelligence, available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence.
[44] European Commission, Ethics guidelines for trustworthy AI, available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[45] European Commission, EU Member States sign up to cooperate on Artificial Intelligence, available at: https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence.

While European countries draw on these EU documents for guidance, there is still significant variation in how directly they integrate them into their strategies. While most reference the Coordinated Plan, human rights and regulation goals are not as actively engaged as, for example, economic benchmarks. As we noted in section 3, this may be because European countries are assuming that their commitment to the EU documents themselves is sufficient representation of their human rights commitments.

Similar to the EU's Coordinated Plan on AI, the Nordic Council of Ministers has a published strategy for the Nordic-Baltic region that establishes shared priorities reflected in the national strategies surveyed. Unlike the EU's Coordinated Plan, the Nordic-Baltic strategy is much less extensive, only establishing a few points for future coordination. Though brief, the strategy pulls out a key trend shared by states in the region: mitigating risks to individual freedoms through ethical governance and protection is argued as the key area through which this region will set itself apart and will have "advantages in international competition."[46]

5.4 The Middle East

The three Middle Eastern strategies developed (from the United Arab Emirates (UAE), Saudi Arabia and Qatar) share goals and guidelines, providing some limited evidence for regional trends in the Middle East. The strategies focused chiefly on data governance and highlight the importance of pro-business regulation regarding data sharing and portability as a central goal. Central to the strategies is making the country attractive for business and leaders in the development of AI technology.

The Saudi Arabia strategy exemplifies these goals through utilizing third party indexes as benchmarks of the strategy's success. Their goals include becoming a Top 10 country in the Open Data Index and being within the Top 20 countries in peer reviewed Data & AI publications. The UAE similarly focuses on regional competition in reducing transaction costs through data regulation; they argue their strategy is "the first of its kind in the region and the world". Qatar is no different in this respect; making "Qatar an attractive jurisdiction to incorporate by AI driven businesses around the globe" is a first priority. Qatar stands out as focusing explicitly on the need for transparency and ethical governance, although it also frames solving these rights concerns as serving business interests in addition to mitigating individual risks.

Outside of these countries, Israel is perhaps the most advanced in terms of actions and measures taken in relation to AI governance, although much of the effort is driven toward research, with the establishment of a cross-governmental team in 2020 to devise recommendations for a policy plan to promote AI research and innovation activities in the country.

At the regional level, the League of Arab States has established an Arab Working Group on AI, chaired by Egypt which, among other things, is considering "developing an Arab AI strategy", as well as identifying AI issues and areas of priority to Arab countries, enhancing cooperation to bridge the digital divide between developed and developing countries, and overcoming a disparity of technological capabilities through exchanging expertise and developing a joint framework for capacity building in the Arab World.[47]

[46] Nordic Co-operation, AI in the Nordic-Baltic region, available at: https://www.norden.org/en/declaration/ai-nordic-baltic-region.
[47] Ministry of Communications and Information, Technology, Egypt Elected Chair of Arab AI Working Group, 16 February 2021, available at: https://mcit.gov.eg/en/Media_Center/Latest_News/News/57187.

5.5 The Indo-Pacific

Many of the strategies in the Indo-Pacific region focus on "human-centered" approaches to AI, outlining the concept of "making AI work for humans" as a key goal. Most of the strategies nod to potential ethical challenges raised by the development of AI but see the development of ethical guidelines and inquiry as outside of the explicit scope of the overall strategy, instead establishing a separate working group or strategy for this focus.

Some of the East Asian strategies highlight the need to articulate the purpose of AI beyond simply encouraging the development of AI as a central concern. Singapore emphasizes focusing on "benefits to citizens and businesses, i.e. getting AI to serve human needs, rather than developing the technology for its own sake".[48] Japan and South Korea similarly focus on a human-centered approach in their wide-spread development plans. South Korea stands out as identifying the need to "raise public awareness of the negative impacts of technological innovation."[49]

Australia and India engage most deeply with ethical risks and regulatory challenges but establish the development and articulation of ethical principles as separate goals from the main overarching strategy. Australia published their ethical principles in a separate supplementary document to their overarching strategy, and India's strategy proposes new research centers to develop policy recommendations for how to balance privacy and ethics with security concerns.

Outside of NASs, Thailand's Office of the National Digital Economy and Society Commission is in the process of developing ethical guidelines for the use of AI with a draft being considered by the country's Cabinet.

5.6 Africa

Mauritius is the only country in Africa to have adopted a comprehensive NAS. The document makes minimal reference to human rights, however, simply noting that "It is clearly visible that the use of AI has implications for privacy, data protection and the rights of individuals". While Egypt announced that it had adopted an NAS in 2020, it is not a comprehensive strategy as such but a short vision and mission statement, alongside a list of strategy pillars, strategy enablers and five priority sectors (agriculture/environment and water management, healthcare, Natural Language Processing, economic planning, and manufacturing and infrastructure management). The document does not make any reference to human rights, although it does note the implementation of the strategy will be monitored and that this will include consideration of laws and regulations, ethical principles and guidelines.

In the absence of strategies, a growing number of governments are considering AI through other policy documents. In 2019, for example, South African President Ramaphosa created a multistakeholder Presidential Commission on the Fourth Industrial Revolution with a mandate to develop an integrated strategy and plan to respond to the Fourth Industrial Revolution. The focus of the Commission's work was on how to achieve global competitiveness in key economic sectors such as agriculture, mining and finance. As such, the report of the Commission is strongly skewed towards recommendations to encourage investment, research and

[48] Smart Nation Singapore, National Artificial Intelligence Strategy, available at: https://www.smartnation.gov.sg/docs/default-source/default-document-library/national-ai-strategy-summary.pdf?sfvrsn=55179e0f_4.
[49] Government of the Republic of Korea, Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution, available at: https://k-erc.eu/wp-content/uploads/2017/12/Master-Plan-for-the-intelligent-information-society.pdf.

development,[50] with consideration of human rights absent both from the Commission's mandate and report. In Zambia, the government's Smart Zambia e-Government Master Plan (2018 – 2030), yet to be adopted, will likely set out how the government plans to use AI within government.

The approach of establishing an independent commission to advise the government has also been taken up in two other countries. In 2018, the Tunisian Secretary for Higher Education and Scientific Research announced the creation of a Task Force to oversee the development of a National Artificial Intelligence Strategy alongside a Steering Committee to devise a methodology and an action plan for its production.[51] While details are scarce, the government did state that one of its aims was to enhance sustainable and equitable development and, while not referring to human rights, stated that ethical challenges would be considered. In Kenya, the government established a Distributed Ledgers Technology and Artificial Intelligence Taskforce to develop a comprehensive strategy to encourage and adopt these emerging technologies. The Task Force's terms of reference made clear that its focus was to make recommendations which would help deliver the government's Big Four Agenda of affordable housing, universal healthcare, manufacturing and agriculture processing, as well as to promote and enhance government services. Despite this, human rights do receive some mention in the report's recommendations. The report notes, for example, the potential impacts that the technologies have on the right to privacy, and recommends that policies that enable both short and long-term use of AI should ensure that citizens' rights are protected. The report also highlights the general risk of "unethical AI applications". Similar task forces, commissions and working groups have also been established in Uganda (in 2019, focusing on the Fourth Industrial Revolution) and, in Nigeria, 2020 saw the establishment of the National Agency for Research in Robotics and Artificial Intelligence.

At a regional level, the African Union established a working group on Artificial Intelligence at the end of 2019 in order to study "the creation of a common African stance on Artificial Intelligence", "the development of an Africa wide capacity building framework" and the "establishment of an AI think tank to assess and recommend projects to collaborate on in line with Agenda 2063 and the UN Sustainable Development Goals". As of the time of writing, no further details of the work of this working group had been made available save that it held its first meeting in December 2019. The SMART Africa Initiative, comprising over 30 African states and with a mission to "accelerate sustainable socio-economic development on the continent, ushering Africa into a knowledge economy through affordable access to Broadband and usage of Information and Communications Technologies" has announced a project supported by the German Development Cooperation to strengthen local technical knowledge on AI and to support the development of AI policy frameworks across the continent. Again, as of the time of writing, no further details of this project had been made available.

[50] Commission on the Fourth Industrial Revolution, Summary Report and Recommendations, December 2019.

[51] As of March 2021, the government of Tunisia appears to be the only the third on the continent, following Mauritius and Egypt, that has committed to developing a National Artificial Intelligence Strategy: Agence Nationale de la Promotion de la Recherche Scientifique, National AI Strategy: Unlocking Tunisia's capabilities potential, 2018, available at: http://www.anpr.tn/national-ai-strategy-unlocking-tunisias-capabilities-potential.

Annex 1

National Artificial Intelligence Strategy human rights assessment tool

1. Methodology / Framework

1.1 The components of an NAS

There is no single standardized structure for an NAS, nor an exhaustive list of the elements that one should contain. As noted in section 1.2.2 of this report, however, there are a small number of taxonomies which suggest the structures and elements that an NAS should use and contain, and there is a significant degree of overlap in the suggested structures and elements contained within key guidance on the development of NASs. In developing the framework set out in the next section, we have reviewed some of the existing taxonomies that have been published:

- The World Economic Forum's "Framework for Developing a National Artificial Intelligence Strategy"[52]
- Dutton's "Building an AI World, Report on National and Regional AI Strategies"[53]
- Dutton's "AI Policy 101: An Introduction to the 10 Key Aspects of AI Policy"[54]

Following a review of these guidance documents and taxonomies, as well as existing NASs, we mapped out their recommended structures for an NAS and identified thirteen core components which are consistently included. These components may not directly correspond to the structures suggested by the guidance documents and taxonomies. Some components are worded differently, grouped together under a broader heading, or divided into more distinct sections. However, together, all thirteen components capture all elements of an NAS recommended by the different guidance documents and taxonomies. In section 1.2.2 of this report, we map out the structures and elements suggested by the guidance documents we reviewed and how they link to the thirteen components that we use.

Those thirteen components are, in summary:

1. **Framing, Vision and Objectives:** This component sets out the government's overall approach towards (or vision of) AI, as well as the aims and objectives of the NAS or the principles which underpin it.
2. **Research:** This component looks at the government's plans to support research in the development and use of AI.
3. **Talent:** This component looks at how the government plans to attract and create domestic and international talent in AI.
4. **Skills and the Future of Work:** This component looks at both how the government plans to ensure that the population, particularly those in education, have the skills needed to engage in a world with AI, as well as plans to help address the disruption in labor as a result of AI.
5. **AI in the Private Sector:** This component sets out the government's plans to promote the use of AI in the private sector.
6. **AI in the Government:** This component sets out the government's plans to promote the use of AI in government and the public sector.
7. **Data:** This component sets out the government's plans to address issues relating to data created by AI, such as data protection frameworks, data sharing arrangements, and open data.

[52] World Economic Forum, *A Framework for Developing a National Artificial Intelligence Strategy: Centre for Fourth Industrial Revolution*, August 2019.
[53] Dutton, T., *Building an AI World, Report on National and Regional AI Strategies*, 2019.
[54] Dutton, T., *AI Policy 101: An Introduction to the 10 Key Aspects of AI Policy*, 2018, available at: https://medium.com/politics-ai/ai-policy-101-what-you-need-to-know-about-ai-policy-163a2bd68d65

8. **Infrastructure and Cybersecurity:** This component sets out the government's plans to ensure that the right physical and other infrastructure is in place for AI, as well as to ensure its security.
9. **Ethics:** This component sets out the government's plans to make sure that the development and use of AI in the state is ethical.
10. **Regulation:** This component sets out the government's plans to introduce or reform legislative and regulatory frameworks relating to AI.
11. **Inclusion:** This component sets out the government's plans to ensure that the benefits of AI are inclusively and equitably felt, as well as to ensure that the AI community broadly is inclusive of diverse backgrounds and perspectives.
12. **Foreign Policy and International Cooperation:** This component sets out how the government will work with other governments, as well as international and regional organizations, on issues related to AI. This usually involves collaboration to deal with shared threats, but may also include promoting the government's particular values and foreign policy priorities at international and regional forums where AI is discussed.
13. **Governance and Stakeholder Engagement:** This component sets out the roles and responsibilities of different actors in the field of AI. This includes those actors which have responsibility for the implementation of the strategy as well as those with whom the government will work or support, such as the private sector or civil society.

While all components have links to human rights, not every section within each component will do; some parts of an NAS will be hugely important from a human rights perspective, while others will have little or no relevance. As such, the recommendations set out in this tool do not touch upon all aspects of an NAS but focus instead on those with the clearest and strongest links to human rights.

1.2 Human rights analysis

As with our report, our analysis and recommendations for this tool are based on existing international human rights law and standards, primarily the Universal Declaration on Human Rights and the International Covenant on Civil and Political Rights, as well as their elaboration and interpretation by the UN Treaty Bodies.[55]

1.3 Scope and rationale for inclusion of good practice examples

As well as setting out criteria for how an NAS can respect, protect and promote human rights, we have included, in the third section of this tool, a number of examples of existing good practice for each of the criteria. In determining which examples to include, we first of all reviewed existing NASs, applying the criteria. From the list of instances where the criteria were met in an NAS, we chose examples which we considered to most strongly meet the criteria, while ensuring as wide a range as possible of regional and national contexts. As such, there are many further instances of good practice which are not highlighted in this tool, and so the examples included should not be considered exhaustive.

[55] While the approach of this tool is based upon international human rights law and standards, other regional human rights systems may provide a greater degree of protection. Governments developing NASs in states which are members of regional human rights system may also wish to consider ensuring that the NAS reflects and is consistent with those regional standards.

1.4 How to use this tool

The second section of this tool contains key questions and criteria for assessment for each of the thirteen components most commonly found in an NAS. The key question(s) are the questions that the drafters of the NAS should ask themselves in relation to each component. These questions should prompt consideration of how human rights relate to that particular component and guide the inclusion of appropriate language and commitments.

While the particular language and commitments will vary from NAS to NAS, reflecting the different contexts of the states which are developing them, each component also contains broad criteria for assessing whether that element sufficiently incorporates human rights considerations. These criteria can be used by the drafters of the NAS themselves as guidance for the types of language and commitments they should include. They can also be used by other stakeholders to assess whether an NAS (or a draft of one) sufficiently incorporates human rights considerations.

As noted above, the third section of this tool provides a number of examples of existing good practice for each of the criteria.

2. Components, Key Questions and Criteria

(1) FRAMING, VISION AND OBJECTIVES

Key Question

- How can human rights considerations be embedded at the heart of the NAS and made central to the strategy?

Criteria for Assessment

- ❑ 1A: The section which frames AI and/or sets out the government's vision of AI should explicitly highlight the links between AI and human rights and the potential benefits and risks to human rights that stem from its development and use.

- ❑ 1B: One of the objectives of the NAS, or one of the principles which underpins it, should be to respect, protect and promote the human rights of persons within the jurisdiction of the state concerned.

(2) RESEARCH

Key Questions

- What steps can be taken to ensure that research into AI adequately respects human rights and that researchers actively work to mitigate human rights risks?
- What can be done to ensure that research will actively promote human rights-respecting applications of AI?

Criteria

- ❑ 2A: The NAS should set out specific actions that the government will take to ensure that any public funding of AI-related research will be human rights-respecting. These steps could include:

    - ❑ Requiring research grant applications to set out how they are human rights-respecting, and for this to be a key consideration in funding decisions.
    - ❑ Requiring prospective research projects to undergo human rights impact assessments in order to identify and mitigate risks to human rights that might materialize as a result of the research.
    - ❑ Specifically placing as a condition for funding evidence that the research will not undermine human rights.
    - ❑ Promoting collaboration between different research disciplines on AI-related funding, building capacity within AI research more broadly on understanding and integrating human rights consideration.

- ❑ 2B: The NAS should set out the steps that the government will take to support research which specifically focuses on the societal impacts of AI and how human rights can be protected in the development and use of AI.

(3) TALENT

Key Question

- What steps can be taken to ensure that specialists in AI have a full understanding of its human rights dimensions to make them better prepared to integrate this into their work?

Criteria

- ❑ 3A: Where this section sets out steps to train AI specialists, it should promote the incorporation of human rights into any such education and training (see also 4A and 9A).

- ❑ 3B: This section should outline specific steps that will be taken to promote a diverse AI workforce with representation from all segments of society.

(4) SKILLS AND THE FUTURE OF WORK

Key Questions

- What steps can be taken to ensure that the future workforce who develop or use AI have a full understanding of its human rights dimensions?
- What must be done to ensure that the right to work and the right to just and favorable conditions of work are protected as AI impacts upon labor markets and employment?

Criteria

- ❑ 4A: Where this section sets out steps to train students, workers or others more broadly on AI, it should promote the incorporation of human rights into any such education and training (see also 3A and 9A).

- ❑ 4B: The NAS should set out the specific steps that the government will take to address threats to the right to work posed by AI, such as the disruption of labor markets.

- ❑ 4C: The NAS should set out the specific steps that the government will take to address threats to the right to the enjoyment of just and favorable conditions of work (including safe working conditions, a workplace free of discrimination and harassment) posed by certain applications of AI, such AI-based surveillance of performance tracking by employers in the workplace.

(5) AI IN THE PRIVATE SECTOR

Key Question

- What steps can be taken at the national level to ensure that human rights are not undermined by the development and use of AI in the private sector?

Criteria

- ❏ 5A: The NAS should set out steps to ensure that appropriate measures, including regulatory measures, are in place to ensure that harms to human rights are prohibited in the private sector (see also 9A).

- ❏ 5B: The NAS should commit the government to ensuring that individuals who are impacted by AI or other algorithmic tools being used to make decisions in the private sector which affect their human rights or legal rights are able to challenge those decisions (see also 9A).

- ❏ 5C: The NAS should commit the government to ensuring that individuals who are impacted by AI or other algorithmic tools being used to make decisions in the private sector which cause adverse human rights impacts have access to grievance and remedial processes, with effective remedies, whether through state-based, company-based or other mechanisms.

(6) AI IN THE GOVERNMENT

Key Question

- What steps must be taken to ensure that the use of AI by the government and the public sector does not undermine human rights?

Criteria

- ❏ 6A: The NAS should commit the government to undertaking human rights impact assessments prior to any AI development, commissioning, procurement or use by all levels of government or in the public sector in order to identify and mitigate risks to human rights that might materialize as a result of the research. Further, these assessments should continue thereafter at all stages of development and implementation of AI in the public sector.[56]

- ❏ 6B: The NAS should commit the government to full transparency over the development, commissioning or use of AI by the government or in the public sector, so that individuals are fully aware of when and how AI is being used (see also 9A).

- ❏ 6C: The NAS should commit the government to ensuring that individuals who have decisions made by AI in the public sector which affect their human rights or legal rights are able to challenge those decisions (see also 9A).

- ❏ 6D: The NAS should commit the government to ensuring that individuals who are impacted by AI or other algorithmic tools being used to make decisions by government or the public sector which impermissibly interfere with their human rights have access to grievance and remedial processes with effective remedies.

[56] See, for example, the Canadian government's Directive on Automated Decision Making, available at: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

(7) DATA

Key Question

- What steps can be taken to ensure that the right to privacy and other human rights are not undermined by the collection, storage and processing of data by AI or in the development of new AI technologies?

Criteria

- ❑ 7A: The NAS should recognize that the use of data by AI represents potential risks to individuals' rights to privacy (including through data collection and storage, as well as the use of data in training models) and set out the steps that the government will take to ensure that these risks are mitigated.

- ❑ 7B: Where the NAS sets out the steps that the government will take to encourage greater open data or more open datasets, this section should also set out the steps that the government will take to ensure that any risks to privacy are mitigated.

- ❑ 7C: As one of these steps, the NAS should commit the government to the development of, where it does not already exist, appropriate, proportionate and effective data protection legislation, consistent with Council of Europe Convention 108, OECD Guidelines on Privacy, and other international best practice (such as the EU's General Data Protection Regulation).

- ❑ 7D: The NAS should recognize the risks of bias and discrimination that stem from use of non-representative or skewed data sets by AI and set out the steps that the government will take to ensure that these risks are mitigated.

(8) INFRASTRUCTURE AND CYBERSECURITY

Key Question

- What is necessary to ensure that measures to protect infrastructure and cybersecurity are not taken at the expense of human rights?

Criteria

- ❑ 8A: The NAS should make clear that any steps taken to enhance the cybersecurity of relevant digital infrastructure will be consistent with the state's obligations under international human rights law, particularly the right to privacy. In doing so, the NAS should clearly outline the tradeoffs that it anticipates in attempting to ensure security.

(9) ETHICS

Key Questions

- What are the specific principles that must be outlined to ensure that human rights are respected at all levels of the development and implementation of AI?
- How will the government ensure that the development and use of AI, regardless of by whom or for what purpose, does not put human rights at risk?

Criteria

- ❑ 9A: The NAS should be specific about the principles that the state is using to ensure the ethical implementation of AI, beginning with how human rights principles apply to and should be considered in the context of AI.

- ❑ 9B: The NAS should set out the specific steps that the government will take to ensure that AI is only developed and used in ways which are human rights-respecting. These steps could include:

    - ❑ Developing a comprehensive legislative or regulatory framework on the development and use of AI, applying to both the public and private sector, which sets out constraints to ensure that harms to human rights are prohibited (see also 10A).
    - ❑ Working with relevant parts of the private sector to ensure that appropriate measures are in place to ensure that harms to human rights are prohibited in the sector (see also 5A).
    - ❑ Requiring greater transparency whenever AI is used in the private and public sector (see also 5A and 6A).
    - ❑ Promoting the incorporation of human rights into the education and training of engineers, researchers and others involved in the development of AI (see also 3A and 4A)
    - ❑ Mechanism for auditing AI and algorithms specifically.

(10) REGULATION

Key Questions

- Are legislative or regulatory measures needed to ensure that safeguards for human rights are effective and enforceable?
- If so, what types of regulation are necessary, and how can risks to human rights from these regulations be mitigated?

Criteria

- ❑ 10A: Where the government is considering developing regulation of AI, the NAS should commit the government to undertaking a comprehensive human rights impact assessment of any legislation, to full public consultation of any legislative proposals, and to the involvement of all relevant stakeholders in the development of any legislation (see also 9B).

(11) INCLUSION

Key Questions

- What steps must be taken to ensure that the rights to equality and nondiscrimination are not undermined by the development and use of AI?
- What steps must be taken to ensure that neither the benefits nor the risks from AI are concentrated in only certain groups of society?
- What can be done to ensure that the development and use of AI involves a diverse and representative group, reflecting society more broadly?

Criteria

- ❑ 11A: The NAS should map out the different stakeholder groups who need to be involved in discussions around AI, particularly in the governance and implementation of the NAS, and set out the steps that the government will take to bolster their inclusion.

- ❑ 11B: The NAS should map out the different communities and social groups who are at risk of being left behind as AI is developed and used and set out the steps that the government will take to mitigate these risks.

- ❑ 11C: The NAS should map out the different communities and social groups who may be particularly adversely affected by AI in different ways and set out the steps that the government will take to bolster their inclusion in AI-related discussions.

(12) FOREIGN POLICY

Key Questions

- What steps can be taken to ensure that international and regional standards and policies relating to AI are also consistent with human rights standards?
- How can the state's foreign policy positions and engagement encourage this?

Criteria

- ❑ 12A: There should be an unambiguous commitment to promoting a human rights-respecting approach to the development and use of AI as part of the state's foreign policy.

- ❑ 12B: The section should identify relevant international and regional forums and policymaking spaces where co-operation on AI takes place and where that foreign policy can be advanced.

- ❑ 12C: The section should set out a process for developing guidelines or guidance on the sale and export of AI technologies in order to ensure that they are not used in ways that might violate human rights.

- ❑ 12D: The government should commit to helping partner countries build AI capacity in a human rights-respecting way and tie cooperation to partner countries' commitment to this approach.

(13) GOVERNANCE AND STAKEHOLDER ENGAGEMENT

Key Questions

- What is the best way to ensure that the full range of relevant stakeholders are engaged in the governance and oversight of the NAS and its implementation?

Criteria

❏ 13A: Governance and implementation of the NAS should involve representatives of all stakeholder groups, including the private sector, the technical community and civil society.

3. Good Practice Examples

(1) FRAMING, VISION AND OBJECTIVES

- ❑ 1A: The section which frames AI and/or sets out the government's vision of AI should explicitly highlight the links between AI and human rights and the potential benefits and risks to human rights that stem from its development and use.

Belgium

"The development, deployment, and use of AI must happen, but with caution for individual and collective rights. AI can indeed pose trust and ethical questions. For instance, AI often uses data based on human behaviour and, hence, can reinforce unwanted human bias. Multiple incidents have been shared in the press over the last few years: an algorithm mistakenly classifying black people as "gorillas" or a recruiting tool favouring men for technical jobs. Moreover, through their scalable nature, algorithms can create negative feedback loops and, in that way, deeply influence our society. For example, police presence based on historical crime can cause a further detection of petty crimes in problematic areas. These same petty crimes would go unnoticed in historically safer areas, further increasing inequality."

Denmark

"Artificial intelligence entails an entirely new way of making decisions, in which computers and algorithms play a larger role. For example, the introduction of self-learning algorithms on social media creates uncertainty regarding the extent to which we can rely on recommendations from intelligent systems.

This raises a number of questions regarding responsibility and security that need to be addressed. Furthermore, artificial intelligence raises a number of ethical issues relating to the relationship between, on the one hand the advantages from using new technologies, and, on other hand, consideration of people's basic rights, due process, and fundamental social values."

Malta

"AI raises profound questions across ethical, legal and regulatory domains, touching a range of areas from protecting national security and citizen rights to advancing commercial interests and international standing. These include the risks of biased and unaccountable automated decision-making, discrimination, data privacy-related issues, cyber threats and the potential for manipulation of political systems and wider society in general."

The Netherlands

"Artificial intelligence (AI) is radically changing the world. AI will make a substantial contribution to economic growth, prosperity and well-being of the Netherlands. It will also be of huge assistance in dealing with societal issues in areas such as ageing, climate change, food safety and healthcare. At the same time, we must not close our eyes to challenges such as the protection of fundamental rights including privacy, non-discrimination and autonomy."

Norway

"Norwegian society is characterised by trust and respect for fundamental values such as human rights and privacy. The Government wants Norway to lead the way in developing and using AI with respect for individual rights and freedoms. This can become a key advantage in today's global competition.

The Government believes that:

- artificial intelligence that is developed and used in Norway should be built on ethical principles and respect human rights and democracy
- research, development and use of artificial intelligence in Norway should promote responsible and trustworthy AI
- development and use of AI in Norway should safeguard the integrity and privacy of the individual
- cyber security should be built into the development, operation and administration of systems that use AI
- supervisory authorities should oversee that AI systems in their areas of supervision are operated in accordance with the principles for responsible and trustworthy use of AI"

❑ 1B: One of the objectives of the NAS, or one of the principles which underpins it, should be to respect, protect and promote the human rights of persons within the jurisdiction of the state concerned.

Czech Republic

"We are going to focus on protecting every person and consumer, their rights and privacy, especially the weakest ones. We are going to prevent discrimination, manipulation and misuse of AI, we are going to set the rules for decision-making of algorithms about people in everyday life."

Denmark

"The goals of the government are that:

- Ethical principles are incorporated in the development and use of artificial intelligence to secure respect for individuals and their rights, and for democracy."

Germany

"Thirdly, the Strategy is based on the democratic desire to anchor such a far-reaching technology as AI, which may also be deployed in sensitive areas of life, in an ethical, legal, cultural and institutional context which upholds fundamental social values and individual rights and ensures that the technology serves society and individuals."

Lithuania

"Having the capability to generate tremendous benefits for individuals and society, AI also gives rise to certain risks that should be properly managed. Given that, on the whole, AI's benefits outweigh its risks, we must ensure to follow the road that maximizes the benefits of AI while minimizing its risks. To ensure that we stay on the right track, a human-centric

approach to AI is needed. Trustworthy AI has two components: (1) ethical purpose - it should respect fundamental rights, applicable regulation and core principles and values and (2) it should be technically robust and reliable since, even with good intentions, a lack of technological mastery can cause unintentional harm."

The Netherlands

"Track 3: Strengthening the foundations: Public values and human rights remain protected"

Russia

"The basic principles of the development and use of artificial intelligence technologies, the observance of which is obligatory during the implementation of this Strategy, include:

a) the protection of human rights and liberties: ensuring the protection of the human rights and liberties guaranteed by Russian and international laws, including the right to work, and affording individuals the opportunity to obtain the knowledge and acquire the skills needed in order to successfully adapt to the conditions of a digital economy;"

(2) RESEARCH

- ❑ 2A: The NAS should set out specific actions that the government will take to ensure that any public funding of AI-related research will be human rights-respecting. These steps could include:
    - ❑ Requiring research grant applications to set out how they are human rights-respecting and for this to be a key consideration in funding decisions.
    - ❑ Requiring prospective research projects to undergo human rights impact assessments in order to identify and mitigate risks to human rights that might materialize as a result of the research.
    - ❑ Specifically placing as a condition for funding evidence that the research will not undermine human rights.
    - ❑ Promoting collaboration between different research disciplines on AI-related funding, building capacity within AI research more broadly on understanding and integrating human rights consideration.

- ❑ 2B: The NAS should set out the steps that the government will take to support research which specifically focuses on the societal impacts of AI and how human rights can be protected in the development and use of AI.

Czech Republic

"Legal and societal aspects of AI, ethical rules, consumer protection and security issues

Long-term objectives (until 2035):

- Securing standards primarily in the areas of security, personal data protection and the protection of fundamental rights in research, development and use of AI."

(3) TALENT

- ❑ 3A: Where this section sets out steps to train AI specialists, it should promote the incorporation of human rights into any such education and training (see also 4A and 9A).

- ❑ 3B: This section should outline specific steps that will be taken to promote a diverse AI workforce with representation from all segments of society.

(4) SKILLS AND THE FUTURE OF WORK

- ❑ 4A: Where this section sets out steps to train students, workers or others more broadly on AI, it should promote the incorporation of human rights into any such education and training (see also 3A and 9A).

- ❑ 4B: The NAS should set out the specific steps that the government will take to address threats to the right to work posed by AI, such as the disruption of labor markets.

Belgium

"We, therefore, must ensure that all higher education students receive at least a soft skilling, data, technology and AI courses in any field, which also covers the ethical aspects."

Norway

"Work on privacy by design and ethics require those who work on solutions based on AI to possess or acquire the necessary competence. Higher education institutions ought to evaluate how privacy and ethics can be integrated into their programmes in, for example, information technology and data science."

- ❑ 4B: The NAS should set out the specific steps that the government will take to address threats to the right to work posed by AI, such as the disruption of labor markets.

Australia (Australia's Tech Future)

"The impact of technological change on current and future jobs is likely to be uneven. Different industries and regions will be impacted in different ways. Many Australians fear ongoing technological change means that they will be left unemployed or underemployed without the skills required to secure one of the new well-paid jobs on offer. Governments and industry need to provide support for workers needing to up-skill, re-skill or transition into new areas of employment, whether this be early in their career or when the person is closer to retirement."

- ❑ 4C: The NAS should set out the specific steps that the government will take to address threats to the right to the enjoyment of just and favorable conditions of work (including safe working conditions, a workplace free of discrimination and harassment) posed by certain applications of AI, such AI-based surveillance of performance tracking by employers in the workplace.

(5) AI IN THE PRIVATE SECTOR

- ❑ 5A: The NAS should set out steps to ensure that appropriate measures, including regulatory measures, are in place to ensure that harms to human rights are prohibited in the private sector (see also 9A).

- ❑ 5B: The NAS should commit the government to ensuring that individuals who are impacted by AI or other algorithmic tools being used to make decisions in the private sector which affect their human rights or legal rights are able to challenge those decisions (see also 9A).

- ❑ 5C: The NAS should commit the government to ensuring that individuals who are impacted by AI or other algorithmic tools being used to make decisions in the private sector which cause adverse human rights impacts have access to grievance and remedial processes, with effective remedies, whether through state-based, company-based or other mechanisms.

(6) AI IN THE GOVERNMENT

- ❑ 6A: The NAS should commit the government to undertaking human rights impacts assessments prior to any AI development, commissioning, procurement or use by all levels of government or in the public sector in order to identify and mitigate risks to human rights that might materialize as a result of the research. Further, these assessments should continue thereafter at all stages of development and implementation of AI in the public sector.[57]

- ❑ 6B: The NAS should commit the government to full transparency over the development, commissioning or use of AI by the government or in the public sector, so that individuals are fully aware of when and how AI is being used (see also 9A).

- ❑ 6C: The NAS should commit the government to ensuring that individuals who have decisions made by AI in the public sector which affect their human rights or legal rights are able to challenge those decisions (see also 9A).

- ❑ 6D: The NAS should commit the government to ensuring that individuals who are impacted by AI or other algorithmic tools being used to make decisions by government or the public sector which impermissibly interfere with their human rights have access to grievance and remedial processes, with effective remedies.

(7) DATA

- ❑ 7A: The NAS should recognize that the use of data by AI represents potential risks to individuals' rights to privacy (including through data collection and storage, as well as the use of data in training models) and set out the steps that the government will take to ensure that these risks are mitigated.

[57] See, for example, the Canadian government's Directive on Automated Decision Making, available at: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

Australia (Artificial Intelligence: Solving problems, growing the economy and improving our quality of life)

"AI developers also need certainty and guidance around what represents ethical and acceptable use of private data. This requires attention to "data creep"; the gradual increase in the comprehensiveness and granularity of data held by organisations about people which, when cross-referenced to other data, provides even more detailed personal insights. Organisations will increasingly be challenged with achieving data integrity whereby people's private and confidential information is properly protected and managed."

Germany

"In order to achieve the goals set out in this strategy, the quantity of useful, high-quality data must be significantly increased without violating personal rights, the right to control one's own data or other fundamental rights."

"In order for high-quality AI applications to be developed, which will prove the prowess of German and European AI developers and users, AI applications require high-quality datasets. In cases where personal data is used, it also needs to be ensured that the data is processed in a way that conforms with the law – i.e. legal quality – by respecting personal rights, the right to control one's own data and other fundamental values. With the EU General Data Protection Regulation having entered into force in May 2018, the EU has adopted high data protection standards, providing a universal legal framework governing the processing of personal data within the EU. The Federal Government will explore as to whether the current legal framework will allow personal data to be used as an opportunity for economic development on the one hand, and ensure the right to control one's own data on the other, taking into account new AI-based technologies."

- ❏ 7B: Where the NAS sets out the steps that the government will take to encourage greater open data, or more open datasets, this section should also set out the steps that the government will take to ensure that any risks to privacy are mitigated.

Germany

"In order to achieve the goals set out in this strategy, the quantity of useful, high-quality data must be significantly increased without violating personal rights, the right to control one's own data or other fundamental rights."

- ❏ 7C: As one of these steps, the NAS should commit the government to the development of, where it does not already exist, appropriate, proportionate and effective data protection legislation, consistent with Council of Europe Convention 108, OECD Guidelines on Privacy, and other international best practice (such as the EU's General Data Protection Regulation).

- ❏ 7D: The NAS should recognize the risks of bias and discrimination that stem from use of non-representative or skewed data sets by AI and set out the steps that the government will take to ensure that these risks are mitigated.

(8) INFRASTRUCTURE AND CYBERSECURITY

- ❑ 8A: The NAS should make clear that any steps taken to enhance the cybersecurity of relevant digital infrastructure will be consistent with the state's obligations under international human rights law, particularly the right to privacy. In doing so, the NAS should clearly outline the tradeoffs that it anticipates in attempting to ensure security.

(9) ETHICS

- ❑ 9A: The NAS should be specific about the principles that the state is using to ensure the ethical implementation of AI, beginning with how human rights principles apply to and should be considered in the context of AI.

- ❑ 9B: The NAS should set out the specific steps that the government will take to ensure that AI is only developed and used in ways which are human rights-respecting. These steps could include:

    - ❑ Developing a comprehensive legislative or regulatory framework on the development and use of AI, applying to both the public and private sector, which sets out constraints to ensure that harms to human rights are prohibited (see also 10A).
    - ❑ Working with relevant parts of the private sector to ensure that appropriate measures are in place to ensure that harms to human rights are prohibited in the sector (see also 5A).
    - ❑ Requiring greater transparency whenever AI is used in the private and public sector (see also 5A and 6A).
    - ❑ Promoting the incorporation of human rights into the education and training of engineers, researchers and others involved in the development of AI (see also 3A and 4A).
    - ❑ Mechanism for auditing AI and algorithms specifically.

Colombia (translated from Spanish)

"Ethical framework for AI and security: the national government recognises that the use of AI entails a series of ethical challenges that must be considered and addressed by the state, such as, justice, freedom, non-discrimination, transparency, responsible design, security, privacy and the role of human rights, among many others.

These principles should be discussed and constructed with the support of the private sector and the scientific and academic community of the country expert in the field.

First, the Ministry of Information Technology and Communications together with the Presidency of the Republic will design a transversal ethical framework that will guide the design, development, implementation and evaluation of AI systems that are implemented in the country, following the principles of the OECD and as indicated in the principles for the development of AI in Colombia, which this document deals with. This framework will also differentiate and emphasize data ethics, algorithm ethics and ethics that guide the behaviors and practices of individuals who develop and implement technology. The proposed principles will develop, at a minimum, concepts such as justice, transparency, freedom, responsibility, inclusion and the role of human rights in the creation and implementation of this technology."

Czech Republic

"Legal and societal aspects of AI, ethical rules, consumer protection and security issues

Long-term objectives (until 2035):

- Securing standards primarily in the areas of security, personal data protection and the protection of fundamental rights in research, development and use of AI."

European Union

"As a first step to address ethical concerns, draft AI ethics guidelines will be developed by the end of the year, with due regard to the Charter of Fundamental Rights of the European Union."

Lithuania

"Establish AI ethics committee that reviews impact of technology on fundamental rights. The committee should include representatives from academia, government, industry and NGO sector. The committee should provide (independently created) short and long-term analysis and recommendations. These recommendations should be used to create and update existing ethical standards."

Luxembourg

"In order to ensure that legal and ethical guidelines are implemented to protect fundamental rights and freedoms, Luxembourg will focus on the following key actions:

- Engaging with the national data protection authority and leveraging its expertise in order to address AI-related questions.
- Setting up a governmental technology & ethics advisory committee to discuss ethical implementations of technology and advising the Government on potential risks and societal impacts.
- Collaborating with key bodies that work on developing and safeguarding corporate governance in Luxembourg to accelerate the adoption of proper AI corporate governance.
- Actively following, under the leadership of ILNAS, international normalization processes in the field of AI, for example in the context of ISO.
- Developing and implementing innovative privacy-enhancing technologies for the use of large datasets in the context of AI learning."

The Netherlands

"The Netherlands is committed to the development and application of responsible AI, which means that the AI must benefit people, that fundamental (European) human rights are protected and that we strive to ensure that everyone is included."

"Actions:

- Commissioned by the government, the Scientific Council for Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid, WRR) will investigate the impact of AI on public values.
- Commissioned by the Research and Documentation Centre (Wetenschappelijk Onderzoek- en Documentatiecentrum) (part of the Ministry of Justice and Security),

Utrecht University has started a research programme under the name 'Legal aspects of decision-making algorithms' (Juridische aspecten van algoritmen die besluiten nemen). This exploratory study involves a study of five cases to identify the key opportunities and risks of decision-making algorithms over the next five to ten years and how they relate to the existing legal frameworks (and the values that lie behind them). The cases to be examined are: the self-driving car, P2P energy markets, judges, 'doenvermogen' (self-efficacy) and content moderation on platforms.

- On 7 June 2019, the Minister for Legal Protection sent a letter to the House of Representatives about the protection of horizontal privacy (or 'social privacy'). The letter contains measures that are partly related to the risks that the use of AI may entail for horizontal privacy.
- On behalf of the Minister for Legal Protection, Tilburg University is conducting research into the risks to our privacy associated with the use of facial recognition technology and into possible measures to limit these risks. This study is expected to be completed by the end of 2019.
- The Minister for Legal Protection is examining the desirability of establishing a system of certification of AI applications across all disciplines in the administration of justice, with the certification bodies being accredited by a body at EU level.
- The Ministry of the Interior and Kingdom Relations is setting up a transparency lab for government organisations, where knowledge is exchanged and support is provided in the areas of transparency, explainability and accountability.
- Together with the Ministry of Justice and Security, Statistics Netherlands (CBS), the Directorate-General for Public Works and Water Management (Rijkswaterstaat) and the Association of Netherlands Municipalities (VNG), the Ministry of the Interior and Kingdom Relations is mapping out the considerations for government organisations that play a role in whether or not algorithms should be made public."

Norway

"The Government will:

- encourage development and use of artificial intelligence in Norway to be based on ethical principles and to respect human rights and democracy
- encourage industry and interest organisations to establish their own industry standards or labelling or certification schemes based on the principles for responsible use of artificial intelligence
- encourage the educational institutions to consider how privacy and ethics can be given a central place in their programmes in artificial intelligence
- expect the supervisory authorities to have the competence and authority to supervise artificial intelligence systems within their areas of supervision in order to, among other things, ensure compliance with the principles for responsible and trustworthy artificial intelligence
- establish a cooperation forum for consumer, competition and data protection enforcement bodies: Digital Clearinghouse Norway
- continue to participate in European and international forums, including the EU's work towards creating a regulatory framework to promote responsible and trustworthy use of artificial intelligence and towards modernising consumer rights in light of digital developments
- stimulate public debate on the ethical use of artificial intelligence"

(10) REGULATION

- ❏ 10A: Where the government is considering developing regulation of AI, the NAS should commit the government to undertaking a comprehensive human rights impact assessment of any legislation, to full public consultation of any legislative proposals, and to the involvement of all relevant stakeholders in the development of any legislation (see also 9B).

(11) INCLUSION

- ❏ 11A: The NAS should map out the different stakeholder groups who need to be involved in discussions around AI, particularly in the governance and implementation of the NAS, and set out the steps that the government will take to bolster their inclusion.

- ❏ 11B: The NAS should map out the different communities and social groups who are at risk of being left behind as AI is developed and used and set out the steps that the government will take to mitigate these risks.

Australia (Australia Tech Future)

"Addressing barriers to digital literacy and access to technology is key to ensuring participation in the economy and for social inclusion, particularly for:

- older Australians
- women
- Indigenous Australians
- people with disabilities
- people in low socio-economic groups
- people living in regional and remote areas."

- ❏ 11C: The NAS should map out the different communities and social groups who may be particularly adversely affected by AI in different ways and set out the steps that the government will take to bolster their inclusion in AI-related discussions.

(12) FOREIGN POLICY

- ❏ 12A: There should be an unambiguous commitment to promoting a human rights-respecting approach to the development and use of AI as part of the state's foreign policy.

Germany

"One area where AI has a beneficial or reinforcing effect is the implementation of the sustainable development goals (SDGs) set out under the United Nations Agenda 2030, as AI technology can help boost the efficiency of energy installations, improve medical diagnosis methods and enhance design measures geared towards adapting to climate change. Potential risks include the use of AI technology in way that violates personal freedoms, the right to control one's own data, privacy and data protection, the use of loopholes in security systems (cyber-attacks) and various forms of discrimination."

European Union

"The EU can make a unique contribution to the worldwide debate on AI based on its values and fundamental rights."

- ❑ 12B: The section should identify relevant international and regional forums and policymaking spaces where co-operation on AI takes place and where that foreign policy can be advanced.

- ❑ 12C: The section should set out a process for developing guidelines or guidance on the sale and export of AI technologies in order to ensure that they are not used in ways that might violate human rights.

- ❑ 12D: The government should commit to helping partner countries build AI capacity in a human rights-respecting way and tie cooperation to partner countries' commitment to this approach.

Germany

"When it comes to developing common guidelines, we advocate taking a multilateral approach by using existing forums such as the OECD, G7, G20 and the United Nations."

(13) GOVERNANCE AND STAKEHOLDER ENGAGEMENT

- ❑ 13A: Governance and implementation of the NAS should involve representatives of all stakeholder groups, including the private sector, the technical community and civil society.

Czech Republic

- "Convening stakeholder working groups to coordinate individual chapters to meet their objectives and keeping the AI Committee updated on their functioning."
- "Cooperation with private and non-State actors institutionalized in the form of memoranda."

Annex 2

Full list of National Artificial Intelligence Strategies (and accompanying ethics documents) reviewed

Australia: Australian Government and CSIRO's Data61, Artificial Intelligence: Solving problems, growing the economy and improving our quality of life, 2019

Australia: Australian Government, Australia's Tech Future: Delivering a strong, safe and inclusive digital economy, 2018

Australia: Australian Government, Department of Industry, Science, Energy and Resources, AI Ethics Principles, 2019

Austria: Federal Ministry of Transport, Innovation and Technology and Federal Ministry of Digital and Economic Affairs, Artificial Intelligence Mission Austria 2030, 2019

Belgium: AI for Belgium, 2019

Canada: CIFAR, Pan-Canadian Artificial Intelligence Strategy, 2017

China: State Council of China, "New Generation Artificial Intelligence Development Plan", 2017

Colombia: National Planning Department, Ministry of Information and Communications Technologies, and Administrative Department of the Presidency of the Republic, National Policy for Digital Transformation and Artificial Intelligence, 2019

Czech Republic: Ministry of Industry and Trade of the Czech Republic, National Artificial Intelligence Strategy of the Czech Republic, 2019

Denmark: The Danish Government, Ministry of Finance and Ministry of Industry, Business and Financial Affairs, National Strategy for Artificial Intelligence, 2019

Egypt: Ministry of Communications and Information Technology, National AI Strategy, 2020

Estonia: Government of the Republic of Estonia, Estonia's national artificial intelligence strategy 2019-2021, 2019

European Union: European Commission, Artificial Intelligence for Europe, 2018

Finland: Ministry of Economic Affairs and Employment, Finland's Age of Artificial Intelligence: Turning Finland into a leading country in the application of artificial intelligence, 2017

Finland: Ministry of Economic Affairs and Employment, Leading the way into the age of artificial Intelligence, 2019

France: Villani, C., For a Meaningful Artificial Intelligence: Towards a French and European Strategy, 2019

Germany: The Federal Government, Artificial Intelligence Strategy, 2018

Hungary: Ministry for Innovation and Technology, Hungary's Artificial Intelligence Strategy 2020-2030, 2020

India: NITI Aayog, National Strategy for Artificial Intelligence #AIForAll, 2018

Italy: Agenzia per l'Italia Digitale, AI White Paper, 2018

Japan: Strategic Council for AI Technology, Artificial Intelligence Technology Strategy, 2017

Lithuania: Ministry of Economy and Innovation, Lithuanian Artificial Intelligence Strategy: A Vision of the Future, 2019

Luxembourg: The Government of the Grand Duchy of Luxembourg, Artificial Intelligence: a strategic vision for Luxembourg, 2019

Malta: Office of the Prime Minister, Parliamentary Secretary for Financial Services, Digital Economy and Innovation, Malta: The Ultimate AI Launchpad: A Strategy and Vision for Artificial Intelligence in Malta 2030, 2019

Malta: Office of the Prime Minister, Parliamentary Secretary for Financial Services, Digital Economy and Innovation, Malta: Towards Trustworthy AI, 2019

Mauritius: Mauritius Artificial Intelligence Strategy, 2018

The Netherlands: Strategic Action Plan for Artificial Intelligence, 2019

Nordic-Baltic Region: Nordic Council of Ministers, AI in the Nordic-Baltic region, 2018

Norway: Ministry of Local Government and Modernisation, National Strategy for Artificial Intelligence, 2020

Portugal: AI Portugal 2030, 2019

Qatar: Qatar Center for Artificial Intelligence, National Artificial Intelligence Strategy for Qatar, 2019

Russia: Decree of the President of the Russian Federation No. 490 of 10 October 2019 On the Development of Artificial Intelligence in the Russian Federation

Saudi Arabia: Saudi Data & AI Authority: National Strategy for Data & AI, 2020

Serbia: Strategy for the Development of Artificial Intelligence in the Republic of Serbia for the period 2020-2025, 2020

Singapore: Smart Nation Singapore, National Artificial Intelligence Strategy: Advancing our Smart Nation Journey, 2019

South Korea: Government of the Republic of Korea, Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution, 2017

Spain: Government of Spain, National Artificial Intelligence Strategy, 2020

Sweden: Government Offices of Sweden, National approach to artificial intelligence, 2018

Taiwan: Taiwan AI Action Plan, 2018

United Arab Emirates: UAE Strategy for Artificial Intelligence, 2017

United Kingdom: HM Government, Industrial Strategy: Artificial Intelligence Sector Deal, 2017

United States: Executive Office of the President, Executive Order 13859 of February 11, 2019, Maintaining American Leadership in Artificial Intelligence

Uruguay: Presidency of the Republic of Uruguay, Artificial Intelligence Strategy for Digital Government, 2020

GLOBAL PARTNERS DIGITAL

Second Home
68 Hanbury St
London E1 5JL

+44 203 818 3258