



**SUSTAINABLE
DEVELOPMENT GOALS**

RESOURCE GUIDE ON ARTIFICIAL INTELLIGENCE STRATEGIES

Resource Guide

ON

ARTIFICIAL INTELLIGENCE (AI) STRATEGIES

.....

April 2021



United Nations

Department of
Economic and
Social Affairs



United Nations
Educational, Scientific and
Cultural Organization



WFEO / FMOI



United Nations
Office of the Secretary-General's
Envoy on Technology

Table of Contents

Chapter 1: Introduction	4
Chapter 2: Ethics of AI	6
Introduction	6
Ethical implications of AI	8
Call for an ethical approach to development and use of AI	10
United Nations work on AI Ethics	10
Key References	11
International Organizations	11
National Governments	17
Private Organizations	23
Civil Society	29
Chapter 3: Technical Standards for AI and International Strategy	41
Introduction to Technical Standards	41
Standards Development	42
Conformity Assessment	43
Standards in Governance	43
Artificial Intelligence Standards	44
Open Platforms Advancing Various Aspects of AI and Machine Learning	44
AI Standards Strategies	45
Key References	47
Chapter 4: National AI Strategies and International Organizations	56
Introduction	56
Key References	57
National Strategies	57
International Organizations	67
Chapter 5: Observations on the Ways Forward	75
Bibliography	79
Annex I: Roadmap for the UNESCO Recommendation on the Ethics of Artificial Intelligence	84
Annex II: Past, ongoing and future initiatives related, either directly or indirectly, to the ethical, legal and social implications of AI within the UN system	9696
Annex III: Summary Table of Ethical Principles	100100
Annex IV: Artificial Intelligence Terminology	117

Acknowledgement

The Resource Guide on Artificial Intelligence (AI) Strategies is a UN publication laying out existing resources on artificial intelligence ethics, policies and strategies on national, regional and international level. The work on the Guide was led by Wei Liu (DESA) with input from Richard A. Roehl (DESA), under supervision of Shantanu Mukherjee (DESA).

The Guide represents a collaborative effort and reflects the input and contributions of experts in the field of technology and innovation. Overall comments and inputs were received from (in alphabetical order) Joe Hironaka, Maksim Karliuk, Prateek Sibal, Rachel Pollack and Shiyao Ke from UNESCO; Huadong Guo from Chinese Academy of Science; Mario Cervantes, Karin Perset (OECD); Monika Matusiak and Veerle Vandeweerd (European Commission); Naoto Kanehira (World Bank); William Colglazier (the American Association for Advancement of Science); Xiaolan Fu (Oxford University); Yu Ping Chan (United Nations Office of the Envoy on Technology) and Zhengzhong Xu (Chinese Academy of Governance).

Chapter 2 received contributions primarily from UNESCO: Dafna Feinholz, Jo Hironaka, Xianhong Hu, Misako Ito, Melissa Tay Ru Jein, Maksim Karliuk, Shiyao Ke, Rachel Pollack, Sasha Rubel, Prateek Sibal, Cedric Wachholz; Alica Daly (WIPO); Bob Bell Jr. and Pilar Fajarnes Garces (UNCTAD); Ewa Staworzynska (ILO); Inese Podgaiska (Association of Nordic Engineers); Jayant Narayan (World Economic Forum); Merve Hickok (Center for AI & Digital Policy); Maria Jose Escobar Silva (Government of Chile); Majid Al Shehry (Saudi Data and AI Authority); Miguel Luengo-Oroz (UN Global Pulse); Olga Cavalli (South School on Internet Governance); Stephan Pattison (Arm Ltd.) and Vanja Skoric (European Center for Not-for-Profit Law ECNL).

Chapter 3 received contributions primarily from Charles Michael Ovink (United Nations Office for Disarmament Affairs); Gong Ke, William Kelly and Pan Li of the World Federation of Engineering Organizations (WFEO) and Preetam Maloor (ITU).

Chapter 4 received contributions primarily from Christina Pombo Rivera (the Inter-American Development Bank); Elisabetta Zuanelli (the Center for Research and Development on E-Content (CReSEC); Friederike Schüür (UN Global Pulse); Sapienza University of Rome) and Yi Zeng from (Chinese Academy of Science).

Research assistance was provided by Adi Gorstein, Catherine Huilin Deng, Kaidi Guo and Naomi Hoffman.

Views expressed in this Resource Guide are of the authors and do not represent official positions of the United Nations or its Member States. Written comments and feedback to this Guide will be most welcome and should be addressed to Wei Liu (liuw@un.org) and Joe Hironaka (j.hironaka@unesco.org).

Chapter 1: Introduction

Artificial Intelligence¹ (AI) has rapidly emerged in a context in which sustainable development has been the overarching goal of the international community. The United Nations has called upon governments to develop national strategies for sustainable development, incorporating policy measures to achieve the 2030 Agenda for Sustainable Development. While AI technologies may support breakthroughs in achieving the Sustainable Development Goals (SDGs), they may also have unanticipated consequences that will exacerbate inequalities and negatively impact individuals, societies, economies and the environment.

AI implementation will need to be supported by a multi-disciplinary review to steer AI in a direction that will respect human rights and human dignity. AI strategies may necessarily include appropriate regulatory and multi-stakeholder frameworks. Policymakers and development practitioners are thus seeking to better understand and address any policy gaps to safeguard transparency, safety, human rights and ethical standards. At the same time, many technological advances are initiated in the private sector and academia. There are vital knowledge and experiences to share among all Science, Technology and Innovation (STI) stakeholders. In this context, recognition of the need for a multi-stakeholder approach will be essential.

The [UN Technology Facilitation Mechanism](#) (TFM) was created by the [Addis Ababa Action Agenda](#) to support the implementation of the SDGs and was launched with the 2030 Agenda on Sustainable Development in September 2015. From the outset, the Division for Sustainable Development Goals DSDG/DESA has been serving as Secretariat for the “Interagency Task Team on Science, Technology and Innovation for the SDGs” ([IATT](#)) and for the Secretary General’s appointed “Group of high-level representatives of scientific community, private sector and civil society” (10-Member Advisory Group) to support the TFM. The two groups mobilize experts from within and outside the UN system for advancing the SDGs through Science, Technology and Innovation (STI) in various contexts. Since 2015, both groups have been coordinated and supported by DESA/DSDG (2015-present), UNEP (2016-17) and UNCTAD (2018-present). Over the years, IATT membership has increased to include 45 UN entities and more than 120 active staff members – an unprecedented level of cooperation on science and technology across the UN.

The TFM facilitates multi-stakeholder collaboration and partnerships through the sharing of information, experiences, best practices, the development of practical guidance, joint

¹ While there is no one single definition of ‘artificial intelligence’ (AI), this publication defines AI as an ensemble of advanced ICTs that enable “machines capable of imitating certain functionalities of human intelligence, including such features as perception, learning, reasoning, problem solving, language interaction, and even producing creative work”. The definition has been proposed by UNESCO’s World Commission on the Ethics of Scientific Knowledge and Technology (COMEST).

activities at the country level, and policy advice for and among the Member States, civil society, the private sector, the scientific community, United Nations entities and other stakeholders.

The TFM comprises four components: the IATT; the 10-Member Group of representatives from civil society, the private sector and the scientific community; the annual Multi-stakeholder Forum on Science, Technology and Innovation for the SDGs (STI Forum); and the TFM online platform as a gateway for information on existing STI initiatives, mechanisms, and programs. The gateway serves as a ‘one-stop-shop’ for information on science, technology and innovation that can contribute to achieving the SDGs, building partnerships and matchmaking.

One important feature over the last four years’ implementation of the TFM is the STI discussions that take place in a multi-stakeholder setting, departing from more traditional UN format. Member States and other key stakeholders highly appreciate this multi-stakeholder approach, through the STI Forum as well as WSIS and the Internet Governance Forum. The main purpose of this Reference Guide on AI Strategies is for the TFM community to respond to the interests of- and concern expressed by the stakeholders in STI Forum regarding AI related issues as well as to provide an overview of recent literature on global strategy to guide the development of AI to Member States. This document is not a technical guidebook on specific AI applications, but rather a “Resource Guide” or “Primer” mainly directed towards the diplomatic community and other policy makers who are involved in setting the global agenda on AI, with the aim of promoting more meaningful deliberations on AI-related resolutions. The diplomatic community is invited to share this Resource Guide with line ministries from their capitals.

This document is a continuation of the work on the STI for SDGs Roadmaps² focusing on one specific area: AI strategy development. Unlike the Guidebook on Roadmaps, this Reference Guide is not about “how to develop AI strategies” but a collection of key references that can provide a global overview of discussions on AI Ethics, technical standards, and examples of national strategies. A subsequent version of the AI Guidebook is planned that will focus on assessment of AI impacts and guiding principles of how to respond. This Reference Guide comprises three main chapters on AI: Ethical Principles and Impacts; Technical Standards and International strategies; and National Strategies.³

² See: <https://sustainabledevelopment.un.org/tfm#roadmaps>

³ Citations for this chapter:

ECOSOC High-level Political Forum on Sustainable Development. (2019, May 29). *Multi-stakeholder forum on science, technology and innovation for the Sustainable Development Goals: summary by the co-chairs*. Retrieved from

https://www.un.org/ga/search/view_doc.asp?symbol=E/HLPF/2019/6&Lang=E

UNDESA. (2019, May 14). Session 1: Emerging Technology Clusters and The Impact Of Rapid Technological Change On The SDGs. *Sustainable Development Knowledge Platform*. Retrieved from

Chapter 2: Ethics of AI

Introduction

As a general-purpose technology, AI has profound implications for individuals, societies, economies, and the environment. To unlock AI's potential to accelerate SDG achievement, while managing risks, it is important to understand how societies are transformed by disruptive technology.

This work needs to be accompanied by ethical reflection. AI applications are not value neutral. They have influenced human-technology relations in both beneficial and harmful ways. In a remarkable example, AI was used to generate some of the earliest alerts about the COVID-19 outbreak⁴ by routinely scanning hundreds or thousands of governmental and media data sources in multiple languages. Yet the evolution of AI already has posed significant new ethical concerns and challenges. It is widely known that AI systems may incorporate biases, due to the data on which they are trained as well as the choices made by developers while designing AI algorithms to interpret this data. Further, AI-based decision-making processes are not always fully explainable and predictable, and thus can be difficult to understand and redress.⁵

Chapter 2 highlights the ethical implications of AI and ways to address them, through the role and work of the UN system.

Ethical implications of AI

AI-based technologies blur the boundary between human subjects and technological objects.⁶ The application of AI not only has societal implications, which can be ethically

<https://sustainabledevelopment.un.org/index.php?page=view&type=20000&nr=5516&menu=2993>.

UNDESA. (n.d.). Technology Facilitation Mechanism Workstream 10: Analytical work on emerging technologies and the SDGs. *Sustainable Development Knowledge Platform*. Retrieved from <https://sustainabledevelopment.un.org/index.php?page=view&type=12&nr=3335>

⁴ Using artificial intelligence to help combat COVID-19, OECD, 23 April 2020, https://read.oecd-ilibrary.org/view/?ref=130_130771-3jtyra9uoh&title=Using-artificial-intelligence-to-help-combat-COVID-19

⁵ The 'Black-Box' problem of AI systems, understood as the opacity in how AI systems make decisions, raises concerns regarding transparency and accountability in automated decision-making. Several solutions, both technical and operational, have been proposed to address transparency in the use of automated decision-making and generating explanations for why the decisions have been taken. For more details on action being taken by governments, the private sector and academia to address the black-box problem of AI, see page 79-81 of the UNESCO report "Steering AI and Advanced ICTs for Knowledge Societies"

⁶ For a more detailed treatment of the subject/object distinction from a philosophy of technology perspective, see UNESCO COMEST report on Robotics Ethics, available here: <https://unesdoc.unesco.org/ark:/48223/pf0000253952>.

evaluated, but also affects the central categories of ethics: the concepts of agency,⁷ responsibility and our value frameworks.

The increasing autonomy of AI systems raises the question of who should bear ethical and/or legal responsibility for outcomes based on AI systems. An example of this is life and death decisions arrived at by autonomous vehicles.

Autonomous decision-making by AI systems may interfere with moral agency and our understanding of it. For instance, online platforms and search companies rely on content personalization for targeted advertising, interpreting data about an individual's prior interests and engagement in order to tailor news feeds and advertisements.

Studies have suggested that the more accurately a recommendation engine pegs our interests, the faster it traps us in an information bubble. One effect is known as 'echo chambers'— AI algorithms that reinforce users' interests through repeated exposure to content based on their prior beliefs. Another effect is 'filter bubbles', which narrows the scope of content users are exposed to. This leads to communicating with conforming viewpoints, further consuming similar content and not being exposed to other ideas. Such 'filter bubbles' and 'echo chambers' may limit an individual's right to access reliable information and to form opinions freely: a necessary foundation for individuals to exercise freedom of expression. This AI business model also may facilitate the spread of disinformation, hate speech or incitement to violence, with dangerous consequences for peaceful and democratic societies. The logic of the social media business models and AI ranking systems has had a harmful impact on the news media, further weakening press freedom and the rights to freedom of expression and access to information.⁸ Similar consequences are seen in the field of culture, where advertainments are individually tailored to the point where they may impede opening the creative horizon.

Another disruptive potential of AI systems is on moral frameworks. Not only do AI systems have societal effects that can be ethically evaluated, but they also affect the very ethical frameworks we use to evaluate them. For instance, AI-powered care robots might change what people value in caregiving and AI-powered teaching systems might affect criteria for good teaching and education.

Digital transformation is happening at a moment where increased inequalities of income and opportunities are at their highest levels in decades⁹, within and between countries. These technologies may be contributing to deepening such inequalities, while a few firms and a handful of countries own the largest share of these technologies. This can change the landscape beyond recognition.

⁷ In general terms, an agent is a being with the capacity to act, and 'agency' denotes the exercise or manifestation of this capacity.

⁸ UNESCO. 2020. *Reporting Facts: Free From Fear or Favour*, <https://unesdoc.unesco.org/ark:/48223/pf0000375061>

⁹ <http://www.oecd.org/social/inequality.htm>

Core elements in the AI ‘life cycle’ will determine who participates and who doesn’t. This aspect is seen in the key issue of data governance: including data collection, data ownership, data sharing and data access. Data is the core ‘raw material’ of AI that allows countries and firms to advance new solutions and products. Data is highly concentrated and there is no shared international governance framework. The skills needed for data mining and the development of models and algorithms are also highly concentrated. There is thus an essential link between the governance of AI and data.

Member States have recognized that a multi-stakeholder AI ethical framework will be needed to address the many ethical challenges of AI. A short list includes the pace of technological innovation on the digital divide; the risks of creating new forms of exclusions; biases embedded within algorithms, including gender biases; the protection of privacy and personal data; the disruption of governance models; the issues of just distribution of benefits and risks; impacts on employment and the future of work; human rights and dignity; transparency, accountability, responsibility; security and risks arising out of dual-use of technology. These and other issues are further addressed in the publications listed at the end of this chapter.

Call for an ethical approach to the development and use of AI

Because of its profound social implications, many organizations and governments are concerned about the ethical implications of AI. This is seen in the large number of reports on AI and ethics, as well as regional and national initiatives that have articulated principles and values to guide the development and use of AI systems.¹⁰ But the practical governance of AI, as a distributed technology, is being spread across numerous institutions, organizations, and private companies. An overall good governance framework for AI thus requires a pluralistic, multidisciplinary, multicultural, and multi-stakeholder approach. Alongside a professional understanding of AI technologies, a deeper ethical reflection is now called for in order to address fundamental questions about what type of future we want for humanity.

United Nations work on addressing ethical implications of AI

The UN Secretary-General has underlined a need to ensure that AI becomes a force for good. The UN systems’ work on ethics of AI builds on a long history of engagement with ethical concerns related to the development and use of information and communication technologies (ICTs).

The international human rights framework formed the basis of the 2003 World Summit on the Information Society’s (WSIS) Geneva Declaration of Principles, stating that “the use of ICTs and content creation should respect human rights and fundamental freedoms of others, including personal privacy, and the right to freedom of thought, conscience,

¹⁰ Annex II provides a links to several global, regional and national initiatives on the ethical implications of AI. Annex I provides summaries of some of these initiatives.

and religion in conformity with relevant international instruments. "The Geneva Plan of Action, adopted at the 2003 WSIS Summit, identified the "Ethical dimensions of the Information Society" as one of eighteen Action Lines in which governments, civil society entities, businesses and international organizations could work together.¹¹

The UN and the individual institutions within its system have been active in addressing challenges related to development, implementation and use of AI.¹²

In August 2018, the UN Special Rapporteur for the promotion and protection of the right to freedom of opinion and expression, David Kaye, released his report on the implications of artificial intelligence (AI) technologies for human rights. The report presents a framework for a human rights-based approach to these new technologies, with a particular focus on freedom of expression and opinion, privacy and non-discrimination.¹³

In May 2019, the UN System Chief Executives Board for Coordination (CEB) adopted a UN system-wide strategic approach and roadmap for supporting capacity development on AI. It outlines an internal plan to support capacity development efforts related to AI technologies, especially for developing countries, with a particular emphasis on the 'bottom billion', in the context of achieving the SDGs. It has a specific commitment to maintain strong ethical and human rights guardrails, ensuring that AI developments do not exceed the capacity to protect society, particularly marginalized, vulnerable and the poorest populations, including women and girls.

The Secretary-General's High-level Panel on Digital Cooperation's 2019 report "The Age of Digital Interdependence" and the Secretary-General's "Roadmap for digital cooperation"¹⁴, launched on 11 June 2020, present actions and recommendations on how the international community could work together to optimize the use of digital technologies and mitigate the risks. Recommendation 3C is specifically devoted to AI and provides a clear ground for a multi-stakeholder cooperation in ensuring that AI becomes a force for good understanding the risks at hand. As outlined in the Roadmap, the Secretary-General has proposed to establish a multi-stakeholder advisory body on global AI cooperation to provide guidance to himself and the international community on AI that is trustworthy, human-rights based, safe and sustainable and promotes peace.

In 2015, UNESCO Member States committed to promoting human rights-based ethical reflection, research and public dialogue on the implications of new and emerging

¹¹ For more information see:

<http://www.itu.int/net/wsis/docs/geneva/official/poa.html#c10>

¹² Annex I provides some past, ongoing and future initiatives related, either directly or indirectly, to the ethical, legal and social implications of AI within the UN system.

¹³ <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx>

¹⁴ <https://www.un.org/en/digital-cooperation-panel/>

technologies and their potential societal impacts, with the adoption of the Internet Universality framework and the associated R.O.A.M principles.¹⁵

In November 2019, UNESCO was given a mandate by its 193 Member States to start the process of elaborating an international standard-setting instrument on the ethics of AI, in the form of a recommendation.¹⁶ The process of elaborating the Recommendation provides an opportunity for Member States to discuss and agree upon principles and recommended policy actions as ethical and human rights guardrails for the design, development and deployment of AI. It also addresses the concerns of developing countries, the benefit of present and future generations, the 2030 Sustainable Development Agenda, gender and racial bias, inequalities between and within countries, and leaving no one behind.

The first draft of the Recommendation draws on the ongoing work of the UN Secretary-General's High-Level Panel on Digital Cooperation, particularly as it relates to its Recommendation 3C on identifying commonalities among the existing set of AI ethics principles. It also links with other related processes and initiatives within the UN system on the ethics of AI, including ITU's AI for Good Global Summit, the Human Rights Council's resolution on "The right to privacy in the digital age", and others.

The draft Recommendation, to be put before UNESCO Member States at the 41st General Conference in 2021, will provide a foundation to support AI ethics-related work across the UN system.¹⁷

¹⁵ An acronym for Human Rights, Openness, Accessibility to all, and Multistakeholder participation

¹⁶ In the context of UNESCO, recommendations are instruments in which "the General Conference formulates principles and norms for the international regulation of any particular question and invites Member States to take whatever legislative or other steps may be required in conformity with the constitutional practice of each State and the nature of the question under consideration to apply the principles and norms aforesaid within their respective territories" (Article 1(b) of UNESCO's Rules of Procedure concerning recommendations to Member States and international conventions covered by the terms of Article IV, paragraph 4, of the Constitution).

¹⁷ Citations for this chapter:

Hu, X., Neupane, B., Echaiz, L. F., Sibal, P., & Rivera Lam, M. (2019). *Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective*. UNESCO Publishing. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000372132>

McDonald, H. (2020, August 04). Home Office to scrap 'racist algorithm' for UK visa applicants. *The Guardian*. Retrieved from <https://www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racist-algorithm-for-uk-visa-applicants>

Microsoft deletes massive face recognition database. (2019, June 07). *BBC News*. Retrieved from <https://www.bbc.com/news/technology-48555149>

New Zealand. (2020, July). *Algorithm Charter for Aotearoa New Zealand*. Retrieved from https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

Schlosser, M. (2019, October 28). Agency. *The Stanford Encyclopedia of Philosophy* (Winter 2019). Retrieved from <https://plato.stanford.edu/entries/agency/>

Key References

The following is a summary of some publications related to AI ethics and principles available online.

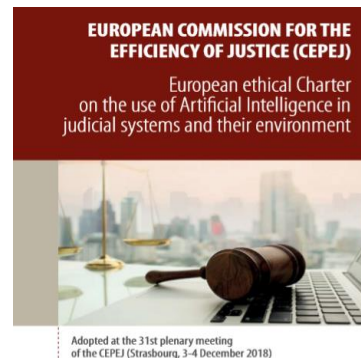
International Organizations

1. European Commission for the Efficiency of Justice: European Ethical Charter on the Use of AI in Judicial Systems

Key words	Developed by	Year
fundamental rights; non-discrimination; security; transparency, impartiality, fairness; user control	CoE CEPEJ	Dec 2018

Acknowledging the potential of AI to improve the efficiency and quality of justice, the charter describes five principles to guide the ethical use of AI specifically in judicial systems, with a focus on processing data and decisions.

The five principles are: respect for fundamental rights; non-discrimination; quality and security (use certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment); transparency, impartiality and fairness; and “under user control” (ensure that users are informed actors and in control of their choice). Each principle is also supported by more concrete recommendations.



Special Address by Antonio Guterres, Secretary-General of the United Nations. (2020, January 23). *World Economic Forum*. Retrieved from <https://www.weforum.org/events/world-economic-forum-annual-meeting-2020/sessions/special-address-by-antonio-guterres-secretary-general-of-the-united-nations-1>

UNESCO. (2020). Records of the General Conference, 40th session, Paris, 12 November-27 November 2019, *volume 1: Resolutions*. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000372579.nameddest=37>

UNSG's High-level Panel on Digital Cooperation. (2019). *The Age of Digital Interdependence*. Retrieved from <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>

Weise, K., & Singer, N. (2020, June 10). Amazon Pauses Police Use of Its Facial Recognition Software. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/06/10/technology/amazon-facial-recognition-backlash.html>

World Commission on the Ethics of Scientific Knowledge and Technology. (2017). *Report of COMEST on robotics ethics*. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000253952>

WSIS. (2003, December 12). Declaration of Principles: Building the Information Society: a Global Challenge in the New Millennium. *World summit on the information society*. Retrieved from <https://www.itu.int/net/wsis/docs/geneva/official/dop.html>

The document also includes a study on the existing uses of AI in judicial systems, covering questions about limitations; a description of potential uses of AI in European judicial systems; and a checklist for integrating the charter's principles into processing methods.

[Download here](#)

2. European Group on Ethics in Science and New Technologies: Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems

Key words	Developed by	Year
human dignity; autonomy; responsibility; justice, equity, solidarity; democracy; rule of law and accountability; security, safety, bodily and mental integrity; data protection and privacy; sustainability	European Group on Ethics in Science and New Technologies	Mar 2018

This statement highlights key ethical questions and considerations relating to AI and proposes a set of fundamental ethical principles, based on the values laid down in the EU Treaties and the EU Charter of Fundamental Rights, which can guide the development of AI. It adopts a rather different approach from other similar documents, framing its principles heavily in terms of human rights and democratic principles.

The principles within the statement are: respect for human dignity (limit to use of algorithms to affect individuals and right to know and decide whether one is interacting with a machine), autonomy (being able to set one's own standards and live according to them, being able to intervene in autonomous systems); responsibility (AI should only be developed in ways that serve social good); justice, equity, and solidarity (equal access to technologies and their benefits, as well as data collection and surveillance); democracy (decisions about AI should be the result of democratic debate, value pluralism and diversity of opinions must not be jeopardized by technologies); rule of law and accountability (right to redress, liability); security, safety, bodily and mental integrity (physical safety, robustness, emotional safety, especially in fields such as cybersecurity and finance); data protection and privacy (right to be free from technologies that influence personal development and to be free from surveillance); and sustainability (environmental friendliness).

The statement concludes by calling for a common, internationally recognized ethical and legal framework for the design, production, use and governance of artificial intelligence, robotics, and 'autonomous' systems.

[Download here](#)

3. European High-Level Expert Group on AI: Ethics Guidelines for Trustworthy AI

Key words	Developed by	Year
Trustworthy AI; lawful; ethical; robust; human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-	European High-Level Expert Group on AI	Apr 2019

discrimination and fairness; societal and environmental well-being; accountability

The European High-Level Expert Group on AI created the concept of 'Trustworthy AI' in response to a mandate to draft ethics guidelines for AI systems that are human-centric, with a goal of improving human welfare and freedom. The report provides examples of opportunities and critical concerns raised by AI, and emphasizes that striving towards trustworthy AI concerns not only the trustworthiness of the AI system itself but also all actors and processes that are part of the system's socio-technical context throughout its entire lifecycle. To this end, trustworthy AI should be lawful, respecting all applicable laws and regulations; ethical, respecting ethical principles and values; and robust, both from a technical perspective while taking into account its social environment.

The Guidelines put forward a set of seven requirements that AI systems should meet in order to be deemed trustworthy, accompanied by a specific assessment list to verify the trustworthiness of any AI system. The key requirements are human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; and accountability.

[Download here](#)

4. ICDPPC: Declaration on Ethics and Data Protection in AI

Key words	Developed by	Year
human rights and fairness, accountability, transparency and intelligibility, responsible development and deployment and respect for privacy, empowerment of individuals and opportunities for public engagement, reduction of biases and discrimination; multi-stakeholder	The International Conference of Data Protection and Privacy Commissioners (ICDPPC)	Oct 2018

Written and sponsored by eighteen data protection and privacy commissioners across the world, the declaration fundamentally aims to preserve human rights in the development of AI. Its guiding principles include respect for human rights and fairness, constant monitoring and accountability, transparency and intelligibility, responsible development and deployment and respect for privacy, empowerment of individuals and opportunities for public engagement, reduction of biases and discrimination. It also suggests several concrete mechanisms for achieving each principle.

The declaration concludes that common governance principles on AI must be established on the basis of a multi-stakeholder approach at an international level.

[Download here](#)

5. ILO: World Employment and Social Outlook (WESO) 2021: The role of digital labour platforms in transforming the world of work

Key words	Developed by	Year
Digital platform	ILO	Feb 2021

The report examines how digital labour platforms are transforming the world of work and the implications of that transformation for employers and workers. It draws on the findings of ILO surveys conducted among some 12,000 workers in 100 countries around the world working on freelance, contest-based, competitive programming and microtask platforms, and in the taxi and delivery sectors. It also draws on interviews conducted with representatives of 70 businesses of different types, 16 platform companies and 14 platform worker associations around the world in multiple sectors.

The report calls for international policy dialogue and coordination to ensure regulatory certainty and the applicability of universal labour standard and lays down 15 recommendations on how to ensure that digital labour platforms provide decent work opportunities and fair competition.

[Download here.](#)



6. OECD: OECD Principles on AI

Key words	Developed by	Year
inclusive growth, sustainable development and well-being; rule of law, human rights, democratic values and diversity; transparency and understandability; robust, secure and safe; accountable	OECD	May 2019

The OECD Principles on Artificial Intelligence promote artificial intelligence (AI) that is innovative and trustworthy and that respects human rights and democratic values. They were adopted in May 2019 by member countries when they approved the OECD Council Recommendation on AI.

The OECD AI Principles are the first such principles signed up to by governments. Beyond OECD members, other countries including Argentina, Brazil, Costa Rica, Malta, Peru, Romania and Ukraine have adhered to the AI Principles, with further adherents welcomed. The OECD AI Principles provided the basis for the G20 AI Principles endorsed by Leaders in June 2019.

There are five principles: AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being; AI systems should be designed to

Artificial intelligence > OECD Principles on AI

What are the OECD Principles on AI?



The OECD Principles on Artificial Intelligence provide a guide to help AI that is innovative and trustworthy and that respects human rights and democratic values. They were adopted in May 2019 by OECD member countries and their governments. The OECD AI Principles are the first such principles signed up to by governments. Beyond OECD members, other countries including Argentina, Brazil, Costa Rica, Malta, Peru, Romania and Ukraine have adhered to the AI Principles, with further adherents welcomed. The OECD AI Principles provided the basis for the G20 AI Principles endorsed by Leaders in June 2019.

In June 2019, the G20 adopted human-centred AI Principles that draw from the OECD AI Principles.

respect rule of law, human rights, democratic values and diversity; there should be transparency and responsible disclosure to ensure understandability; AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed; and organizations and individuals developing, deploying or operating AI systems should be held accountable. The OECD also provides five recommendations to governments on: facilitating investment in AI research and development; fostering a digital ecosystem for AI data, compute and knowledge; shaping an enabling policy environment for AI; building human capacity and preparing for labour market transformation; and co-operating internationally in a multi-stakeholder manner.

[Download here](#)

7.The Institute for Ethical AI: Recruitment AI has a Disability Problem: questions employers should be asking to ensure fairness in recruitment

Key words	Developed by	Year
Fairness; accessibility; inclusion	The Institute for Ethical AI	Sep 2020

This publication advocates for fairness in Recruitment AI for people with disabilities. Fairness in Recruitment have thus far received little attention as Artificial Intelligence (AI) technologies have the potential to dramatically impact the lives and life chances of people with disabilities seeking employment and throughout their career progression.

This white paper details the impacts to and concerns of disabled employment seekers using AI systems for recruitment and provides recommendations on the steps employers can take to ensure innovation in recruitment is also fair to all users. In doing so, it shows that making systems fairer for disabled employment seekers ensures systems are fairer for all.

[Download here](#)

8. United Nations: Secretary-General's Roadmap for Digital Cooperation

Key words	Developed by	Year
Digitalization; inclusive growth; sustainable development; human rights; global cooperation	United- Nations	May 2020



The *Roadmap for Digital Cooperation* is the result of a multi-year, multi-stakeholder, global effort to address a range of issues related to the Internet, artificial intelligence, and other digital technologies, and was presented by the UN Secretary-General on 11 June 2020. The Roadmap builds on recommendations made by the High-level Panel on Digital Cooperation in their June 2019 report.

To meet the call to connect, respect, and protect the online world, the action-oriented Roadmap presents the Secretary-General's recommendations for concrete action by diverse stakeholders that would enhance global digital cooperation in 8 key areas: (1) Achieving universal

connectivity by 2030; (2) Promoting digital public goods to unlock a more equitable world; (3) Ensuring digital inclusion for all, including the most vulnerable; (4) Strengthening digital capacity building; (5) Ensuring the protection of human rights in the digital era; (6) Supporting global cooperation on artificial intelligence; (7) Promoting digital trust and security; (8) Building a more effective architecture for digital cooperation.

[Download here](#)

9. UNESCO: Culture, Platforms and Machines: The Impact of AI on the Diversity of Cultural Expressions

Key words	Developed by	Year
non-biased and non-discriminatory; gender equality; transparent and explainable; auditable and accountable	UNESCO	Nov 2018

The report to the Intergovernmental Committee for the Protection and Promotion of the Diversity of Cultural Expressions reaffirms the importance of an ethical framework for AI and highlights the role of the creative sector in understanding how this framework should look like. The ultimate objective, the report suggests, is that AI systems should be socially beneficial.

To this end, the report echoes existing documents that selection of data and design of algorithms should be non-biased and non-discriminatory; promote gender equality; and be as transparent and explainable as possible. Their creators should also be concretely auditable and held accountable.

The report is unique in its focus on fostering inclusivity and diversity of expression, emphasizing that AI should not homogenize cultural expressions but rather should be used to provide better access to varied expressions and promote perspectives of traditionally marginalized groups.

[Download here](#)



National Governments¹⁸

1. Australia Department of Industry, Innovation and Science: Australia's Ethics Framework

Key words	Developed by	Year
Human, social and environmental well-being; human-centered values; human rights; diversity; autonomy; fairness; inclusive; accessible; discrimination; privacy protection and security; reliability and safety; transparency and explainability; contestability; accountability	Department of Industry Innovation and Science	Nov 2019

The ethics principles were developed following public consultation on a discussion paper. The principles are meant to be aspirational and used together with existing AI-related regulations. It is suggested that the principles can help anyone designing, developing, integrating or using AI to achieve better outcomes; reduce the risk of negative impact; and practice the highest standards of ethical business and good governance, but the framing of the principles seems to target AI developers and business users. It is noteworthy that contestability is noted as a principle in its own right as other frameworks frequently include it under accountability.

The principles are: human, social and environmental wellbeing; human-centered values (human rights, diversity, autonomy); fairness (inclusive, accessible, no unfair discrimination); privacy protection and security (privacy rights, data protection, data security); reliability and safety (operate in accordance with their intended purpose); transparency and explainability; contestability (timely process to allow people to challenge the use or output of the AI system); and accountability (human responsibility and oversight).

[Download here](#)

2. Chinese National Governance Committee for AI: Governance Principles for a New Generation of AI

Key words	Developed by	Year
harmony and friendliness; fairness and justice; inclusivity and sharing; open and orderly competition; privacy; secure/safe and controllable; shared responsibility; open collaboration; agile governance	Chinese National Governance Committee for AI	Jun 2019

¹⁸ The Following list is a summary of known examples and may not be fully comprehensive.

healthy development of a new generation of AI; better coordinate the relationship between development and governance, ensure that AI is safe/secure, reliable, and controllable; promote economically, socially, and ecologically sustainable development; and jointly build a community of common destiny for humanity". The principles are very comprehensive and is similar to the Beijing AI Principles.

The principles are: harmony and friendliness (enhance common well-being of humanity, conform to human values, promote human-machine harmony, safeguard societal security and respect human rights, prohibit malicious application); fairness and justice (eliminate bias and discrimination, promote equality of opportunity); inclusivity and sharing (green development, coordinated development for disadvantaged groups and regions, strengthen AI education, avoid monopolies, open and orderly competition); respect privacy (right to know and right to choose, redress mechanisms); secure/safe and controllable (transparency, explainability, reliability, and controllability, auditability, supervisability, traceability, and trustworthiness, robustness); shared responsibility (adhere to laws, regulations, ethics, standards and norms, accountability, consider risks and impacts); open collaboration (exchanges across disciplines and regions; launch international cooperation; broad consensus on international AI governance framework); and agile governance (ethical development of AI while not hindering innovation, research potential future risks and ensure that AI moves in a direction beneficial to society).

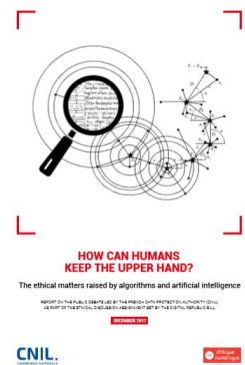
[Download original text here](#)

[\(Official text can be downloaded here\).](#)

3. French Data Protection Authority: How Can Humans Keep the Upper Hand? Report on the Ethical Matters Raised by AI Algorithms

Key words	Developed by	Year
fairness; continued attention and vigilance; requirement for human intervention; intelligibility, transparency and accountability	French Data Protection Authority (CNIL)	Dec 2017

The report is the result of public debate organized by the Authority, involving over 60 partners. The debate identified six main ethical issues, derives several guiding principles, and concludes with practical policy recommendations. The ethical challenges identified are: threat to autonomy and free will; bias, discrimination and exclusion; diminishing collective principles which are the basis of our societies; collection and retention of personal data; which and how much data should be used; and hybridization of humans and machines, of which the threat to collective principles and hybridization are unique.



Two foundational principles and two engineering principles are identified: fairness (should not generate or aggravate any form of discrimination, even if unintentional); continued attention and vigilance (not only to specific applications but at a systemic level); reconsidering the requirement for human intervention (ensure that multi-stakeholders human deliberation governs and guides the use of algorithms and its effects); and intelligibility, transparency and accountability.

[Download here](#)

4. Japan Government: Social Principles of Human-Centric AI

Key words	Developed by	Year
dignity; inclusion and diversity; sustainability; human-centric; human rights; expand human abilities; happiness; prevent overreliance and malicious use; human autonomy and control; user-friendliness; education; eliminate disparities; privacy; accuracy and legitimacy; security; risk management; fair competition; fairness, accountability and transparency; discrimination; dialogue; trust; innovation; collaboration; quality and reliability; accessible data	Government of Japan	Apr 2019

Released by the Cabinet Office, these principles follow the publication of Japan's AI strategy. The principles, together with individual organizations' principles of AI development and utilization, are structured as the foundation of a pyramid, supporting the vision of an AI-ready society and ultimately a philosophy of dignity (prevent overreliance, allow humans to demonstrate their capacities), diversity and inclusion, and sustainability (reduce social disparities, build sustainable societies). Most of the principles are broad ethical values, but education; fair competition; and innovation are more policy recommendations.

The social principles of AI are: human-centric (human rights, expand human abilities, pursue happiness, prevent overreliance and malicious use, human autonomy, user-friendliness); education (eliminate disparities in AI literacy, including understanding of bias, fairness and privacy issues); privacy (protect freedom, dignity and equality, accuracy and legitimacy, human control); security (risk management, sustainability); fair competition (no dominant position); fairness, accountability and transparency (no discrimination, appropriate explanations, opportunities for dialogue, mechanism to secure trust in AI); and innovation (collaboration; quality and reliability; accessible data)

[Download here](#)

Social Principles of Human-centric AI (Draft)

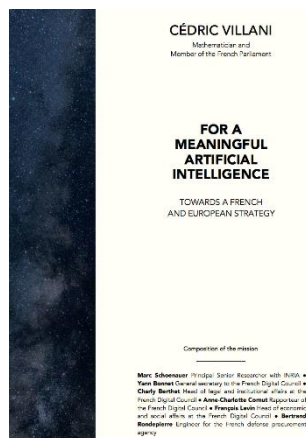
unofficial translation

TABLE OF CONTENTS

1. Introduction	1
2. Fundamental philosophy	3
3. AI Ready Society: Essential social revolution to achieve Society 5.0	5
4. Social Principles of Human-centric AI	6
4. 1. AI Social Principles	6
4. 2. Development and utilization principles of AI	9
5. Conclusions	10

5. Mission Villani: For a Meaningful Artificial Intelligence

Key words	Developed by	Year
transparency and auditability; protection of rights and freedoms; accountability and responsibility; diversity and inclusivity; and political debate transparency and auditability	Mission Villani	Mar 2018



This report lays the foundations for a French strategy in the AI field. The report discusses topics relating to data, the field of AI research, the impact of AI on the economy and employment. Ethics of AI are specifically explored in Part 5 of the document, but Parts 4 and 6 also touch on environmental concerns and inclusivity and diversity, which are considered by many other frameworks to be part of ethical values. It should also be noted that establishing an ethical framework is one of the three main commitments for France (alongside betting on French talent and pooling assets).

Within the specific section on ethics, ethical issues such as explainability, bias and autonomy are discussed and explained with examples. Five principles are suggested as a basis for a future ethical framework: transparency and auditability; protection of rights and freedoms; accountability and responsibility; diversity and inclusivity; and political debate. On the basis of these principles, the strategy suggests opening the black box (ensuring explainability, addressing equity, bias and discrimination, developing auditing systems, research into accountability); considering ethics from the design stage (ethics training for AI researchers, discrimination impact assessment, considering collective rights to data); staying in control (in applications in policing and autonomous weapons); and specific governance of ethics in AI.

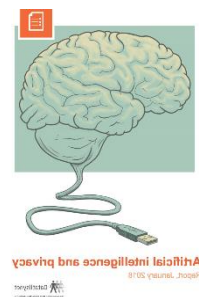
[Download here](#)

6. Norwegian Data Protection Authority: Artificial Intelligence and Privacy

Key words	Developed by	Year
privacy; data protection; fairness; purpose limitation; data minimization; transparent	The Norwegian Data Protection Authority	Jan 2018

The report focuses on challenges in AI relevant to the data protection principles embodied in the General Data Protection Regulation (GDPR). It also highlights how data protection authorities may be able to address any harms caused by AI and offers recommendations to protect these principles for different stakeholders. The target group for this report consists of people who work with, or are interested in, AI.

The four principles identified from the GDPR are: fairness and discrimination; purpose limitation (to whatever user has consented to,



in public interest; data minimization (amount and nature of data used); and transparency and the right to information (and right to explanation).

[Download here](#)

7. Smart Dubai: AI Principles and Ethics

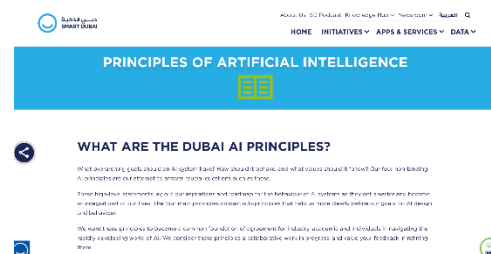
Key words	Developed by	Year
ethics; fair; transparent; accountable; understandable; security; safe; serve and protect humanity; beneficial to humanity; aligned with human values; inclusiveness; global governance; respect dignity and rights.	Smart Dubai ¹⁹	Jan 2019

The Smart Dubai office aims to have these four non-binding, high level statements become a common foundation for industry, academic and individuals navigating the world of AI. Each principle contains sub-principles, each in turn further detailed by operationalizing statements.

The four principles are ethics, making AI systems fair, accountable, as explainable as technically possible, and transparent; security, ensuring that AI systems are safe, secure and controllable by humans, and not able to autonomously hurt, destroy or deceive humans; humanity, planning for a future in which AI systems become increasingly intelligent, and giving AI systems human values and making them beneficial to society; and inclusiveness, promoting human values, freedom and dignity; respecting people's privacy; sharing the benefits of AI throughout society; and governing AI as a global effort.

The ethics principle has also been developed into the Dubai AI Ethics Guidelines, which is accompanied by the Ethical AI Toolkit offers tangible recommendations on how to create AI systems that adhere to the ethics principle.

[Download here](#)



8. Singapore Personal Data Protection: Discussion Paper on Artificial Intelligence

Key words	Developed by	Year
accountability; trust; understanding; explainable; transparent; fair; human-centric; benefice; 'do no harm'	Personal Data Protection Commission Singapore	Jun 2018

¹⁹ Smart Dubai, <https://www.smartdubai.ae/about-us>

This paper presents the Singapore Personal Data Protection Commission (PDPC)'s preliminary analysis of some of the issues pertinent to the commercial development and adoption of AI solutions. The paper proposes an accountability-based framework for discussing ethical, governance and consumer protection issues related to the commercial deployment of AI, particularly for issues relevant to personal data protection. The four-stage framework operationalizes how broad principles can be adopted by stakeholders. It is targeted at the private sector, to encourage them to develop their own voluntary governance frameworks, and stresses that any governance should be 'light-touch' and not prescriptive.



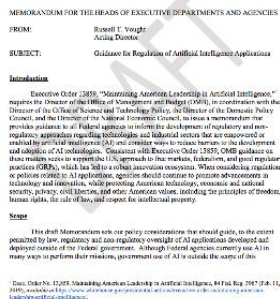
The principles for responsible AI, which aim to promote trust and understanding in AI, are: decisions made by or with the assistance of AI should be explainable, transparent (including accountability) and fair (avoid discrimination); and AI systems, robots and decisions made using AI should be human-centric (confer benefits, should not cause harm, tangible benefits should be identified and communicated, safety).

The four stages of the framework are: identifying the objectives of a governance framework; selecting appropriate organizational governance measures; considering consumer relationship management processes; and building a decision-making and risk assessment framework.

[Download here](#)

9. The White House Office of Science and Technology Policy (OSTP): Principles for the Stewardship of AI Applications

Key words	Developed by	Year
trust; public awareness; integrity; autonomy; flexibility; fairness and non-discrimination; transparency; safety	The White House Office of Science and Technology Policy (OSTP)	Jun 2020



The White House Office of Science and Technology Policy (OSTP), United States, published on January 2020, describes key Principles for the Stewardship of AI Applications. The principles are embedded within a memorandum for the heads of executive departments and agencies on guidance for regulation of AI applications. The memorandum is careful to emphasize the need to encourage innovation and growth by minimizing the regulatory burden, and this attention to minimized regulation is reiterated throughout the description of every principle.

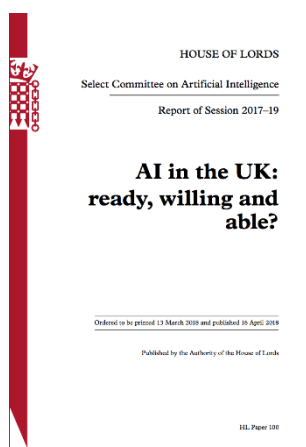
The principles are: public trust in AI (privacy, individual rights, autonomy, civil liberties); public participation and awareness; scientific integrity and information quality (quality, transparency, compliance, bias mitigation, appropriate uses, predictable, reliable and optimized outcomes); risk assessment and management; benefits and costs (full societal costs and benefits, distributional effects, comparison to alternative); flexibility (performance-based, adaptable); fairness and non-discrimination; disclosure and

transparency; safety and security (confidentiality, integrity, availability of information, systemic resilience, preventing malicious use); and interagency coordination (sharing experiences while protecting privacy, liberties and American values).

[Download here](#)

10. UK House of Lords: AI in the UK: ready, willing and able?

Key words	Developed by	Year
common good; benefit of humanity; intelligibility; fairness; data rights and privacy; flourish; autonomous power to hurt, destroy or deceive human beings	UK House of Lords	Apr 2018



This report introduces artificial intelligence, including its definition and history, describes the design and development process for AI, highlighting ethical issues related to data control and transparency, and discusses the impact of AI on various sectors, such as education and healthcare.

The report suggests the development of a core set of widely recognized ethical principles. As a starting point, five overarching principles suggested are: AI should be developed for the common good and benefit of humanity; AI should operate on principles of intelligibility and fairness; AI should not be used to diminish the data rights or privacy of individuals, families or communities; all citizens have the right to be

educated to enable them to flourish mentally, emotionally and economically alongside AI; and the autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence.

The section ends by acknowledging that many other organizations are preparing their own ethical codes of conduct, but that the government should work towards wider awareness and coordination and develop a cross-sector ethical code of conduct with sector-specific variants, an 'AI code', which could provide the basis for statutory regulation if it is deemed necessary.

[Download here](#)

Private Sector

1. ArmBlueprint: Arm AI Trust Manifesto

Key words	Developed by	Year
Security; eliminate discriminatory bias; explainability; human safety	Arm Ltd	Nov 2019

arm Blueprint

The Arm AI Trust Manifesto developed by Arm Ltd, outlines six principles for trust in A.I systems. The main objective of the manifesto

is to ensure that these principles and ethics considerations are incorporated into the key design principles for AI products, services, and components. At present, there is no defining set of practices to follow. Therefore, Arm Ltd calls for an industry-wide effort to define and standardize a set of practices that can be adopted by anyone deploying AI technologies.

The main principles are: AI systems should employ state of the art security; discriminatory bias should be eliminated in designing and developing AI decision systems; technological approaches should be developed to help AI systems record and explain their results; Users of AI systems should have a right to know who is responsible for the consequences of AI decision making; Human safety must be the primary consideration in the design of any AI system; people from all backgrounds should be supported to develop the skills needed for an AI world.

[Download here](#)

2. DeepMind Ethics & Society Principles

Key words	Developed by	Year
privacy, transparency and fairness; AI morality and values; governance and accountability; AI and the world's complex challenges; misuse and unintended consequences; and economic impact: inclusion and equality	DeepMind	Oct 2017

It is important to note that the following are not principles that DeepMind has committed to, but rather research areas that are being explored. These thematic areas are a creation of an independent research unit of DeepMind, a subsidiary of Google's parent company Alphabet Inc. The preamble to the principles highlights that ethical standards and safety as a prerequisite to finding AI's potential benefits, and states DeepMind's belief that AI should be used for socially beneficial purposes and always remain under human control. The thematic areas are: privacy, transparency and fairness; AI morality and values; governance and accountability; AI and the world's complex challenges; misuse and unintended consequences; economic impact and inclusion and equality. The last area's focus on the economic/employment impact is unique among private firms.

[Download here](#)

3. Google: AI at Google: Our Principles

Key words	Developed by	Year
-----------	--------------	------

socially beneficial; public availability of Google information; avoid bias; safety; accountable; privacy; scientific excellence	Jun 2018
---	----------

Google's AI principles are based on the objective of creating technologies that solve important problems and help people in their daily lives. Google's principles set out their commitment to responsible development of technology and identify specific application areas which it will not pursue.

Google commits to developing AI applications based on the following objectives be socially beneficial and facilitate the public availability of high-quality and accurate information while respecting cultural, social and legal norms; avoid creating or reinforcing unfair bias; be built and tested for safety; be accountable to people; incorporate privacy design principles; uphold high standards of scientific excellence; and be made available for particular beneficial uses.

Conversely, Google is unique in explicitly identifying areas where it will not design or deploy AI: technologies likely to cause harm (noting, however, the exception of where "the benefits substantially outweigh the risks); weapons or other technologies whose principal purpose is to cause injury to people; technologies that gather or use information for surveillance violating internationally-accepted norms; and technologies whose purpose contravenes widely accepted principles of international law and human rights.

[Download here](#)

4. IBM: Everyday Ethics for AI

Key words	Developed by	Year
Accountability; value alignment; explainability; fairness; user data rights	IBM	Sep 2018

This document is targeted at designers and developers building and training AI. The five areas of focus are accountability (designers should be responsible for considering the implications of AI systems); value alignment (AI should be aligned with the norms and values of the user group); explainability (easily detectable and decision-making processes are understandable); fairness (minimize bias and promote inclusive representation); and user data rights (protect user data and preserve user power over access and uses). Each area is illustrated using the running example of an AI in-room virtual assistant/concierge for a hotel chain with specified capabilities, and accompanied by concrete recommendations for action, and guiding questions for reflection and future action.

[Download here](#)

5. IBM: Principles for Trust and Transparency

Key words	Developed by	Year
-----------	--------------	------

trust and transparency; augment human intelligence; data and insights belong to their creators; data privacy and security; transparent, explainable and free of bias	IBM	May 2018
--	-----	----------



IBM's core principles for AI are trust and transparency, and include: AI's purpose is to augment human intelligence and not to replace the human workforce; data and insights belong to their creators and data privacy and security is correspondingly respected; and new technology (including its use, purpose, and workings) must be transparent, explainable and free of bias.

IBM's own actions in line with their principles are also listed, and their policy recommendations for governments in light of these principles.

[Download here](#)

6. Microsoft AI Principles

Key words	Developed by	Year
Responsible AI; fairness; reliability and safety; privacy and security; inclusiveness; transparency; and accountability	Microsoft	Nov 2018

Microsoft's AI Principles are encompassed under the term "Responsible AI" and is Microsoft's commitment to AI driven by ethical principles that put people first. The six principles listed are: fairness; reliability and safety; privacy and security; inclusiveness; transparency; and accountability. These are elaborated upon in the book the "Future Computed" and in some short videos.

The Office of Responsible AI (ORA) and the AI, Ethics, and Effects in Engineering and Research (Aether) Committee are responsible for putting these principles into practice. Specific actions taken by Microsoft include applying these principles to their own research and work; helping other organisations develop responsible AI; fostering socially beneficial AI applications; and providing openly-available resources on responsible AI.

[Download here](#)

7. Orange: Human Inside

Key words	Developed by	Year
ethics; human rights; environment; reduce inequalities; responsibility; workplace well-being	Orange	Jan 2019

While Orange does not have a clearly defined framework of AI ethics, it emphasizes the responsible use of AI, guided by its broader philosophy of "Human Inside", which it applies to all of its work. According to this framework, Orange's technical contribution is meant

to be beneficial to individuals, communities and countries, empowering them to take advantage of the digital world, while being environmentally friendly. In AI development specifically, Orange maintains its focus on helping people by making AI a “useful and accessible” innovation that collaborates with man.

[Download here](#)

8. OpenAI Charter

Key words	Developed by	Year
broadly distributed benefits; long-term safety; technical leadership; cooperative orientation	OpenAI	Apr 2018

This charter describes the principles that OpenAI uses to carry out its work, which is narrowly focused on Artificial General Intelligence (AGI), with the broad objective of “acting in the best interests of humanity”. Thus, not all of the principles represented are ethical principles.

The four principles are: broadly distributed benefits for all humanity; long-term safety paying attention to safety precautions; technical leadership; and cooperative orientation through working with others and sharing (safety, policy and standards) research.

[Download here](#)

9. SAP: Guiding Principles for AI

Key words	Developed by	Year
driven by values and laws; inclusive systems that empower humans and augment talents, collaborative and diverse process; reduce bias; transparency and integrity; quality and safety standards; data protection and privacy	SAP	Sep 2018

SAP’s guiding principles for development and deployment of their AI software is designed to “help the world run better and improve people’s lives”. It commits to moving beyond what is legally required and reflect discussions with multiple stakeholders. Unique among the frameworks created by private tech firms, SAP frames its principles as actionable items rather than abstract principles.

The principles are: being driven by values outlined in internal documents and international laws and preventing inappropriate use of their technology; designing inclusive AI systems that seek to empower humans and augment their talents, through a collaborative and diverse process; reduce bias through increasing workforce diversity and investigating new technical methods; striving for transparency and integrity through setting standards, clear communication and client control; upholding quality and safety

standards through testing and working closely with customers, and ensuring data protection and privacy through adherence to regulation and research.

SAP also pledges to engage in debates about wider societal challenges, such as changing nature of skills, role of AI in care, and ethical issues, over which it claims to have less control.

[Download here](#)

10. Tencent Institute: The Ethics of Technology in the Intelligent Age - Reshaping Trust in a Digital Society

Key words	Developed by	Year
trustworthiness; individual well-being; social sustainability; open and inclusive, reliable, understandable and controllable; individual digital well-being, narrowing digital divide and preventing harm; right to fulfilling employment; ability to freely, intelligently, and happily live and develop; inclusive and sustainable development	Tencent Institute	Jul 2019

In this report, China's Tencent group's research institute outlined three dimensions of ethics for socially beneficial technology: trustworthiness of the technology; individual well-being; social sustainability and describes some of Tencent's ongoing initiatives in each dimension.

In the first dimension, technology systems themselves must be available (implying openness and inclusiveness), reliable, understandable and controllable. In the second dimension, technology must co-exist in harmony with humans and facilitate the achievement of personal satisfaction. AI must guarantee individual digital well-being, narrowing the digital divide and preventing harm and misuse; the right to fulfilling employment; and the ability to freely, intelligently, and happily live and develop. In the third dimension, AI should promote the inclusive and sustainable development of the economy, society and healthcare. Its potential to facilitate progress towards 2030 Sustainable Development Goals is particularly highlighted.

[Download here](#)

腾讯发布人工智能伦理报告 倡导面向人工智能的新的技术伦理观

互联网 腾讯科技 2019-07-11 12:50

★ 收藏 0 评论 < 分享

腾讯科技讯 近日，腾讯研究院和腾讯AI Lab联合发布了人工智能伦理报告《智能时代的技术伦理观——重塑数字社会的信任》。人工智能伦理是国际社会近几年关注的焦点话题。最近，国内外都有重大进展，OECD成员国采纳了首个由各国政府签署的AI原则，即“负责任地管理可信AI的原则（Principles for responsible stewardship of trustworthy AI）”，这些原则之后被G20采纳，成为人工智能治理方面的首个政府间国际共识，确立了以人为本的发展理念和敏捷灵活的治理方式。无独有偶，我国新一代人工智能治理原则也于近日发布，提出和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理八项原则，以发展负责任的人工智能。

11. Thales Group: TrUE AI Approach

Key words	Developed by	Year
transparent; understandable; ethical	Thales Group	Jun 2019

Thales' approach to AI was formulated with the French government's AI for Humanity initiative in mind. The main objective of Thales' approach is, therefore, to develop an AI

that “puts the human back in control” through applications of AI that make the world more secure and efficient. To that end, Thales commits to developing AI that is transparent, where users can see the data used to arrive at a conclusion; understandable, that can explain and justify the results and; ethical, that adheres to objective standards protocols, laws, and human rights.

[Download here](#)

Civil Society

1. Amnesty International and Access Now: The Toronto Declaration

Key words	Developed by	Year
human rights; right to equality and non-discrimination; inclusion; diversity; equity; transparency; accountability; multi-stakeholder	Amnesty International and Access Now	May 2018

The Declaration is framed as a way to affirm the existing obligations and responsibilities of both states and private sector actors to promote, protect and respect human rights, as applied to the field of AI. As a corollary, in line with human rights law, it suggests also that use of systems must be transparent, institutions must be held accountable where they fail to protect rights, and also that any discussions surrounding rights should be multi-stakeholder.

While acknowledging that other rights are impacted by AI, the declaration focuses on the right to equality and non-discrimination. To protect this right, the declaration states that all governments and private sector organizations are obligated to prevent and mitigate discrimination risks and ensure adequate remedy in place, and actively promote diversity and inclusion. For each actor, the declaration identifies a list of steps that should be taken for them to be considered in line with the obligation to protect human rights.

[Download here](#)

The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems

Preamble

1. As machine learning systems become increasingly used in society, we must ensure the impact of this technology on human rights. The development and use of machine learning systems must be done in a way that respects human rights. We are concerned about the potential for machine learning systems to discriminate against certain groups of people. We must ensure that machine learning systems do not discriminate against people on the basis of race, ethnicity, gender, age, sexual orientation, disability, or other characteristics. We must ensure that machine learning systems are transparent and accountable. We must ensure that machine learning systems are used in a way that respects human rights.
2. As discussed above, artificial intelligence systems, in this Declaration, refer to those systems that are used to make decisions about individuals or groups of individuals. These systems, which are often used to make decisions about individuals or groups of individuals, have the potential to discriminate against certain groups of people. We must ensure that machine learning systems do not discriminate against people on the basis of race, ethnicity, gender, age, sexual orientation, disability, or other characteristics. We must ensure that machine learning systems are transparent and accountable. We must ensure that machine learning systems are used in a way that respects human rights.
3. This Declaration aims to help to ensure that machine learning systems are used in a way that respects human rights. The goal of this Declaration is to ensure that machine learning systems are used in a way that respects human rights.

1. United Nations Human Rights Council, Human Rights and the Future of Artificial Intelligence, 2018, <https://www.unhcr.org/refugees/2018/04/18-04-2018-human-rights-council-report-on-artificial-intelligence.html>

2. Association of Nordic Engineers: Nordic Engineers’ stand on AI & Ethics

Key words	Developed by	Year
Responsibility; accountability; Transparency; addressing bias; trust; avoiding harm ethics in engineering education	Association of Nordic Engineers (ANE), in cooperation with the IT University of Copenhagen	Sep 2018

On September 25th, 2018, the ANE in cooperation with the IT University of Copenhagen organized an ethics hackathon entitled “Nordic engineers’ stand on the EU future AI and ethics framework”, to gather engineers from five Nordic countries to collaboratively

develop a joint position based on practical experience and in conversation with current debates on AI and ethics. Participants discussed ethical ways of working with AI and considered what they themselves would have liked to have when engaging with developing or implementing AI or what they would like to impart to their junior colleagues. The resulting policy document includes policy recommendations and guidelines of ethical conduct for AI development and implementation.

Engineering plays an essential role in building, sustaining, and improving the quality of life for individuals in contemporary societies. As such, engineers are at the forefront of developing autonomous systems and adding machine intelligence to existing mechanisms and processes. The many standards and codes of conduct agree that one of the major responsibilities of engineers is to promote positive outcomes for society, and to limit harm.

However, in a rapidly changing world, what comes to constitute a positive outcome, and what could potentially cause harm, have become much more difficult to recognize. The existing guidelines and standards often do not fully address the problems that engineers face and the responsibilities they must take on in working with AI. As the stakes rise so does the need for addressing the ethics of engineering in practice more directly.

[Download here](#)

3.Association of Nordic Engineers: Addressing Ethical Dilemmas in AI: Listening to Engineers"

Key words	Developed by	Year
distributed responsibility, accountability, governance, spaces for raising ethical concerns, nodes of certainty, explainability, no-blame culture, oversight.	Association of Nordic Engineers (ANE), the Data Ethics ThinkDoTank (DataEthics.eu), the Institute of Electrical and Electronics Engineers (IEEE) and researchers from the Department of Computer Science at the University of Copenhagen	2021

In September 2020 engineers from across Europe and beyond came together to create the report "Addressing Ethical Dilemmas in AI: Listening to Engineers". The report details engineers' needs and concerns in seeking routes to the development of ethical and responsible AI. The standards and guidelines currently being developed are essential but require processes that ensure their implementation. It is of paramount importance to define specialized responsibilities and put in place local and clearly determined structures for accountability. Principles alone cannot guide our technology development. Engineers stressed the need to better govern decisions about AI technologies in practice.

The main recommendations from the report are: Put in place a governance framework and define and distribute the responsibilities; Engage with stakeholders during system design; Develop and share standards and best practices for addressing ethical challenges;

Create spaces to help engineers identify, discuss and deal with ethical issues; Create new and better technical documentation and certification, and Push for explainability and testing of AI systems before system launch and throughout the system life cycle.

[Download here](#)

4. Association for Computing Machinery: Statement on Algorithmic Transparency and Accountability

Key words	Developed by	Year
transparency; awareness; access and redress; accountability; explanation; data provenance; auditability; validation and testing	Association for Computing Machinery (ACM)	Jan 2017



The statement is premised on the idea that institutions using AI should be held to the same standards of transparency and accountability as institutions using human decision-making, and is consistent with the ACM Code of Ethics. Therefore the 'principles' suggested in the statement are more like criteria for the broader principles of transparency and accountability. They are: awareness of possible biases and their harms; accessibility of algorithmic decisions and redress for harm caused; accountability of institutions which use algorithms; explanation of algorithmic decision-making; clarity of data provenance; auditability; and validation and testing, particularly against discrimination.

[Download here](#)

5. AI Now Institute at New York University: 2019 Report

Key words	Developed by	Year
Fairness, Safety and Security;	AI Now Institute at New York University	2019



The Report highlighted that the spread of algorithmic management technology in the workplace is increasing the power asymmetry between workers and employers. AI threatens not only to disproportionately displace lower-wage earners, but also to reduce wages, job security, and other protections for those who need it most. Efforts to regulate AI systems are underway, but they are being outpaced by government adoption of AI systems to surveil and control. Growing

investment in and development of AI has profound implications in areas ranging from climate change to the rights of healthcare patients to the future of geopolitics and inequities being reinforced in regions in the global South.

[Download here](#)

6. Beijing AI Principles: (Beijing Academy of Artificial Intelligence)

Key words	Developed by	Year
Do good; for humanity; responsible; control risks; ethical; diverse and inclusive; open and share; use wisely and properly; informed consent; education and training; optimizing employment; harmony and cooperation; adaptation and moderation; subdivision and implementation; long-term planning	Beijing Academy of Artificial Intelligence (and officially endorsed by Tsinghua University, Peking University, Chinese Academy of Sciences, and Artificial Intelligence Industry Technology Innovation Strategic Alliance, etc.)	May 2019

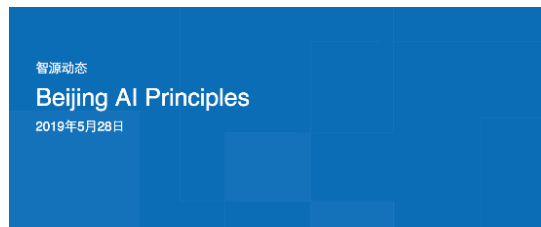
While the AI principles are not officially endorsed or accepted by the Chinese government, the Beijing Academy of AI is backed by the Chinese Ministry of Science and Technology and the Beijing municipal government. The principles call for healthy

development of AI “to support the construction of a human community with a shared future, and the realization of beneficial AI for humankind and nature” and are divided into principles for research and development; for use; and for governance. The principles are extremely comprehensive but reads more as a list of principles, not to position themselves as guidelines for operationalization. Some of the principles also go beyond ethical values, into policy recommendations.

For research and development, the principles are: AI should do good (promote progress of society and human civilization, promote sustainable development); be for humanity (conform to human values and interests such as privacy, dignity, freedom, autonomy, rights, and should not be used against humans); be responsible (consider all risks and take actions to reduce them); control risks (maturity, robustness, reliability, controllability; security and safety); be ethical (trustworthy, fair, reduce discrimination; transparency, explainability, predictability, traceable, auditable, accountable); diverse and inclusive (benefit as many as possible, especially those underrepresented); be open and share (avoid monopolies, equal development opportunities)

For use, the principles are: use wisely and properly (operate according to intended purpose and operators should understand the system); informed consent and redress mechanisms; and education and training (stakeholders should be able to receive education to help them adapt to the impact of AI).

For governance, the principles are: optimizing employment (cautious attitude towards promotion of AI applications that may have impact on employment; explorations on human-AI coordination); harmony and cooperation (avoid AI race, share experience); adaptation and moderation (principles and policies should be actively adjusted to AI development); subdivision and implementation (consider formulating more detailed



guidelines for certain fields of AI applications); long-term planning (research on Artificial General Intelligence and superintelligence should be encouraged).

[Download here](#)

7. Boston Global Forum: AI Social Contract for the AI Age:²⁰

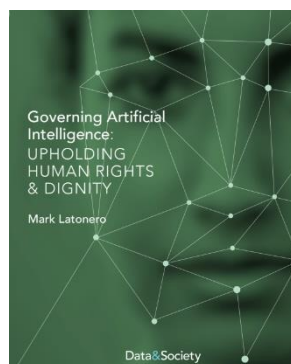
Key words	Developed by	Year
Responsibility; accessibility; integrity; creativity; honesty; tolerance; equality	Boston Global Forum, AIWS, Michael Dukakis Institute, World Leadership Alliance- Club de Madrid	Sep 2020

The Social Contract for the AI Age seeks to build a multi-stakeholder, inclusive society in all aspects of life across politics, government, economics, business, and industry. The Social Contract for the AI Age values creation, innovation, philanthropy, and mutual respect. It seeks the right of freedom on, and access to, the Internet worldwide. The Social Contract for the AI Age seeks to make the world a locus of responsible interaction—a place where every person's contribution is recognized and everyone has a right to knowledge and access to information, where no one is above the law, where money cannot be used to subvert political process, and where integrity, knowledge, creativity, honesty, and tolerance shape decisions and guide policy.

[Download here](#)

8. Data & Society: Governing Artificial Intelligence: Upholding Human Rights & Dignity

Key words	Developed by	Year
international human rights; non-discrimination; equality; political participation; privacy; freedom of expression	Data & Society	Oct 2018



This report proposes the use of a human rights-based framework to provide normative guidance to those developing AI, in order for AI to benefit to the common good, where common good is interpreted as upholding human dignity. The report analyses the impact of AI on five human rights areas through recent news items: non-discrimination, equality, political participation (which in turn implicates the right to self-determination and the right to equal participation in political and public affairs), privacy, freedom of expression, noting that many other human rights are also affected by AI.

The report strongly recommends that the effects of AI on human rights should be constantly monitoring, and that human rights should not be seen as an ethical preference but as fundamental rights that should be enforced through law and regulation and

²⁰ <https://bostonglobalforum.org/highlights/social-contract-for-the-ai-age/>

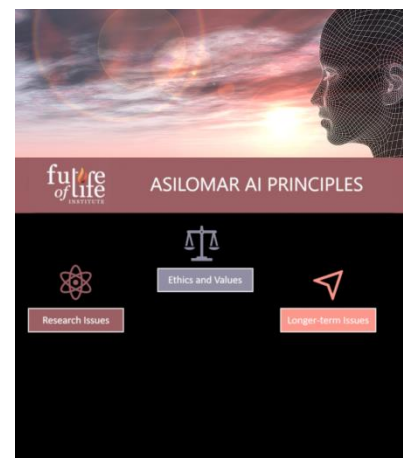
supported by market incentives, public awareness and activities and technological innovation. For technology companies, it suggests that human rights consideration should go beyond statements and be integrated into product and design teams, including in human rights impact assessments, test suites, and product design document. Finally, it acknowledges that human rights laws and principles may not be equipped to address all of the concerns related to AI.

[Download here](#)

9. Future of Life Institute: Asilomar AI Principles

Key words	Developed by	Year
Research goal; research funding; science-policy link; research culture; race avoidance; safety; failure transparency; judicial transparency; responsibility; value alignment; human values; personal privacy; liberty; shared benefit; shared prosperity; human control; non-subversion; AI arms race; capability caution; importance; risks; recursive self-improvement; common good	Future of Life Institute	Jan 2017

This set of 23 principles is one of the leading initiatives calling for a responsible development of AI, having been signed by hundreds of stakeholders, with signatories representing predominantly scientists, AI researchers and industry. Unlike other frameworks, its principles are not limited to abstract ethical values, but also includes within its principles how research and longer-term issues should be guided by ethics. Principles range from broad ethical values to fairly specific directives on particular application areas. Each principle is accompanied by a single sentence of explanation, but not operationalized. The principles are also unique in addressing longer-term issues related to the development of Artificial General Intelligence (AGI). Under research issues, the principles are: research goal should be to create beneficial intelligence; funding should also be directed to research ensuring beneficial use of AI; a strong science-policy link should exist; a culture of cooperation, trust and transparency should be fostered among researchers; and corner-cutting on safety standards should be avoided. For ethics and values, AI systems should be safe; transparent when it fails; transparent when used in judicial decision-making; have responsible stakeholders in designers and builders; align with human values of dignity, rights, freedoms and cultural diversity; respect personal privacy; respect liberty; benefit and empower as many people as possible; create shared prosperity; be under human control; should not subvert social and civic processes; and should not contribute to an arms race in lethal autonomous weapons.



For longer-term issues, we should avoid strong assumptions regarding upper limits on future AI capabilities; advanced AI and its risks should be planned for and managed with care and appropriate resources; AI designed to self-improve or self-replicate must be subject to strict safety and control measures; and superintelligence should only be developed in service of widely-shared ethical ideals and for the benefit of humanity rather than just one state or organization.

[Download here](#)

10. Internet Society: Artificial Intelligence and Machine Learning

Key words	Developed by	Year
Ethics; user-centric; interpretability; public empowerment; responsible deployment; human control; safety; privacy; security; accountability; socioeconomic opportunities; open governance; multi-stakeholder	Internet Society	Apr 2017



Foreword

Artificial intelligence (AI) is a technology that is already impacting how we interact with, and are affected by, the internet. In the near future, its impact is likely to only continue to grow as the potential to vastly change the way that humans interact with one another and the digital world, but also with each other through their work and through other socio-economic institutions – for better or for worse.

It is clear to almost all that the impact of artificial intelligence will be significant, and it will be essential that an open and transparent participation in the debate surrounding AI.

In this paper, we seek to provide an introduction to AI to policymakers and other stakeholders in the wider internet ecosystem.

The paper explains the basics of the technology behind AI, identifies the key considerations and challenges surrounding the technology, and provides several high-level principles and recommendations to follow when dealing with the technology.

If more stakeholders bring their points of view and expertise to the discussions surrounding AI, we are confident that the challenges can be addressed and the best benefits of the technology can be realized.

The paper explains the basics of the technology behind AI, identifies the key considerations and challenges surrounding the technology, and provides several high-level principles and recommendations to follow when dealing with the technology. The paper places ethical considerations (a user-centric approach) as one among other guiding principles and recommendations such as ensuring the “Interpretability” of AI systems; empowering the consumer; responsibility in the deployment of AI systems (human control; safety; privacy; security); ensuring accountability; and creating a social and economic environment that is formed through the open participation of different stakeholders.

[Download here](#)

11. IEEE: Ethically Aligned Design

Key words	Developed by	Year
human rights; well-being; data agency; effectiveness; transparency; accountability; awareness of misuse; and competence	IEEE	Mar 2019

Through an extensive process of public consultation, IEEE formulated its approach to ethically-aligned design, with the ultimate goal for AI systems to remain human-centric, serving humanity's values and ethical principles and benefiting people and the environment, beyond simply reaching functional goals and addressing technical problems. The report summarizes its goal as achieving 'eudaimonia', or human well-being, both at the individual and collective level. Other than listing abstract principles, the report offers scientific/philosophical analysis grounding for the principles and actionable recommendations for standards and regulations. It is targeted at technologists, educators and policymakers. The general principles identified are: protecting and promoting human rights; increased well-being; data agency and control over personal identity; effectiveness and fitness of purpose; transparency of decisions; accountability for decisions; awareness of and guarding against misuse; and competence for safe and effective operation. These principles are categorized under three pillars: universal human values; political self-determination and data agency; and technical dependability. Perhaps due to its nature as a technical organization, IEEE's report is one of the few that highlights effectiveness and competence as principles.

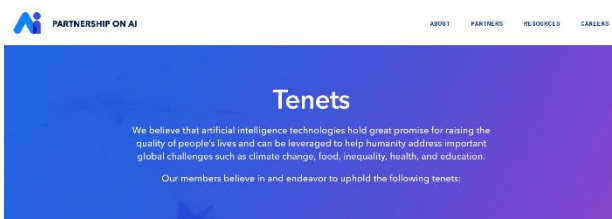


[Download here](#)

12. Partnership on AI: Tenets

Key words	Developed by	Year
Empower; engaging stakeholders; being accountable; representation; privacy and security; respecting interests of all; socially responsible; robust, reliable, trustworthy and secure; human rights; understandable and interpretable by people; cooperation, trust and openness	Partnership on AI	Sep 2016

The Partnership's members believe that AI holds great promise for raising the quality of people's lives and can be leveraged to help humanity address important global challenges such as climate change, food, inequality,



health, and education. The tenets are not a set of abstract ethical principles but rather broad actionable commitments that are underpinned by ethical principles.

To this end, the members commit themselves to: ensuring that AI benefits and empowers as many people as possible; educating, listening to and actively engaging stakeholders to seek their feedback, inform them of their work, and address their questions; open research and dialogue on the ethical, social, economic, and legal implications of AI, engaging with and being accountable to a broad range of stakeholders in their research

and development efforts; engaging with and having representation from the business community; working to maximize benefits and address challenges by protecting privacy and security, understanding and respecting interests of all parties, working to ensure that AI research is socially responsible, sensitive, and engaged directly with its influences on wider society, ensuring that AI is robust, reliable, trustworthy and operates within secure constraints, opposing AI that would violate international conventions or human rights and promoting safeguards; making AI understandable and interpretable by people; and creating a culture of cooperation, trust and openness among AI scientists and engineers. [Download here](#)

13. The Alan Turing Institute: Ethical Platform for the Responsible Delivery of an AI Project

Key words	Developed by	Year
Ethically permissible; fair and non-discriminatory; worthy of public trust; safety, accuracy, reliability, security, robustness; justifiable, transparency, interpretability	The Alan Turing Institute	Jun 2019

These principles are specifically geared towards the design and implementation of AI in the public sector, to manage AI's impacts responsibly and to direct the development of AI systems toward optimal public benefit. It highlights that the consideration of ethics must be incorporated at every stage and involve a collaborative effort between different stakeholders.

The report envisions a three-layered 'ethical platform', or a governance architecture, for an AI-power project in public service. The goal of this 'ethical platform' would essentially be to ensure that the project will uphold the following principles: ethically permissible in terms of the impacts it may have on the well-being of affected stakeholders and communities; fair and non-discriminatory towards all groups; worthy of public trust as it is safe, accurate, reliable, secure and robust; and justifiable through transparency and interpretability of decisions and behaviors.

This platform will be founded upon three building blocks, at different stages of the project. The Support-Underwrite-Motivate (SUM) values of respect, connect, care and protect, aim to provide a framework to consider the societal and ethical impacts of the project and establish criteria to judge its ethical permissibility; the FAST Track principles (fairness, accountability, sustainability, transparency) are a set of actionable principles for responsible design and use; and the process-based governance framework

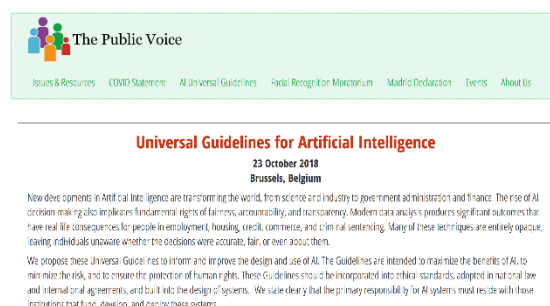


operationalizes the SUM values and FAST Track Principles across the entire project delivery workflow.

[Download here](#)

14. The Public Voice Commission: Universal Guidelines for AI

Key words	Developed by	Year
human rights; transparency; human determination; obligation; identification; fairness; assessment and accountability obligation; accuracy, reliability, and validity obligations; data quality; public safety; cybersecurity; prohibition on secret profiling; prohibition on unitary scoring; termination obligation	The Public Voice Commission	Oct 2018



The Guidelines are meant to maximize the benefits of AI, minimize its risks, and ensure the protection of human rights. The document explicitly states the responsibility for these guidelines lies with the institutions that fund, develop and deploy these systems, and is fairly specific in describing what has to be done/cannot be done in

order to respect these guidelines. It is unique in framing some of its guidelines as 'obligations' and 'prohibitions' rather than principles to follow, lending it a more prescriptive tone than some other frameworks.

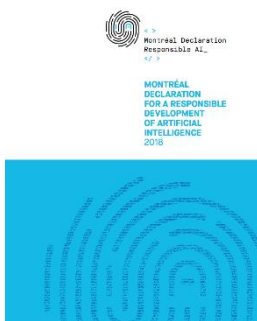
The guidelines are: right to transparency (knowing the basis of AI decision that concerns individuals); right to human determination; identification obligation (institution responsible for an AI system must be made known to the public); fairness obligation (absence of unfair bias and discriminatory decisions); assessment and accountability obligation (systems should only be developed after an evaluation and institutions must be responsible for AI-made decisions); accuracy, reliability, and validity obligations; data quality obligation; public safety obligation; cybersecurity obligation; prohibition on secret profiling; prohibition on unitary scoring (no national government shall establish or maintain a general purpose score on its citizens or residents); and termination obligation (an institution that has established an AI system has an obligation to terminate the system if human control is no longer possible).

[Download here](#)

15. University of Montreal: Montréal Declaration for Responsible Development of Artificial Intelligence

Key words	Developed by	Year
-----------	--------------	------

well-being; respect for autonomy; privacy and intimacy; solidarity; democratic participation; equity; inclusion; prudence; responsibility; sustainable development	University of Montreal	Nov 2017
--	------------------------	----------



The principles outlined in the declaration are premised on the belief that human beings seek to grow as social beings and strive to fulfil their potential by exercising their capacities and aims to ensure the responsible development of AI for the common good in line with this belief. The declaration is unique in having been developed through multi-stakeholder consultation including citizens, experts, public officials, and industry stakeholders. Perhaps as a result, it is unique in including several statements about higher-level subjective well-being requirements, such as not contributing to stress and anxiety; allowing individuals to fulfil their own moral objectives and a conception of a life worth living; integrity of personal identity; fostering human relationships and more.

It is addressed to any entity that wishes to take part in the responsible development of AI in any way and to political representatives who are expected to respond to the risks and opportunities of AI. Within the ethical framework, a few key cross-sectorial themes are covered: algorithmic governance, digital literacy, digital inclusion of diversity and ecological sustainability.

The principles are: growth of the well-being of all sentient beings; respect for people's autonomy and increasing their control over their lives and surroundings; protection of privacy and intimacy; maintaining the bonds of solidarity among people and generations; intelligible, justifiable, accessible and subject to democratic scrutiny, debate and control; contribute to the creation of a just and equitable society; compatible with maintaining social and cultural diversity and must not restrict the scope of lifestyle choices or personal experiences; exercise caution by anticipating adverse consequences and taking appropriate measures to avoid them; must not contribute to lessening the responsibility of human beings in decision-making; and ensure environmental sustainability. Each principle is also accompanied by a few statements which operationalize their meaning. Other than developing an ethical framework for the development and deployment of AI, the declaration also has the objectives of guiding the digital transition so that everyone benefits, and creating a forum for discussion.

[Download here](#)

16. WFEO: Big Data and AI Principles in Engineering

Key words	Developed by	Year
Good for Humanity and Its Environment; Inclusiveness, Fairness, Public Awareness and Empowerment; Opening and Sharing while Respecting Privacy and Data; Integrity; Transparency; Accountability; Peace, Safety and Security; Collaboration	WFEO	Mar 2020

In order to promote responsible conduct of Big Data and Artificial Intelligence (AI) application and innovation in engineering, World Federation of Engineering Organizations (WFEO) has formulated the 7 Principles and releases it now on the first World Engineering Day Celebration.

Engineering societies, as practitioners of Big Data and AI innovation and application, have the responsibility to promote innovation and ensure their development and application to maximize their benefit to people and our living environment while minimizing their negative impact.

[Download here](#)

Chapter 3: Technical Standards for AI and International Strategy

Introduction to Technical Standards

In 2018, the United Nations Economic Commission for Europe (UNECE) co-organized with the International Organization for Standardization (ISO) a conference on the use of voluntary consensus standards in meeting the United Nations Sustainable Development Goals (SDGs) as a side event to the 41st meeting of ISO in Geneva. The conference explored through case histories how standards could be applied to SDG 6 Clean water and sanitation; SDG 7 Affordable and clean energy; SDG 11 Sustainable cities and communities; and SDG 13 Climate action. A number of the case histories provide standardization examples of digitalization of urban services- water and energy and smart cities.

There are many good introductory resources on standardization and many of them are available online. Standards are developed by national, regional and international standards developing organizations and by businesses and other organizations for their own use. Standards are also developed by consortia of businesses to address specific marketplace or industry needs, and by governments to support regulation. The focus in this chapter is on standards developed by ISO and the International Electrotechnical Commission (IEC) and their national members and the International Telecommunications Union (ITU).

IEC and ISO define a standard as "a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context."

No distinction has been made so far between standards in general and technical standards. Technical standards deal with technical systems or the technical aspect of systems and standards more generally with products and processes. Today, voluntary consensus standards are widely used in non-technical areas.

IEC works on electrical and electrotechnical standardization and ISO in virtually all other areas except telecommunications, which is covered by ITU. A joint technical committee (JTC) of ISO and IEC (JTC1) deals with information technology standardization. ISO, IEC, and ITU are all working on standardization related to artificial intelligence (AI) and the "Internet of Things (IoT)."

Standards enable compatibility and interoperability between and among products and systems. They make it easier for consumers and users to compare products. When standards are adopted globally, they facilitate trade and enhance the quality of life globally.

Voluntary consensus standards are just that voluntary. Voluntary standards become mandatory when they are incorporated into business contracts or government regulations. The ISO IEC document Using and referencing ISO and IEC standards to support public policy provides details for policymakers on how to reference ISO and IEC standards and some examples.

There are a number of programs that support standards development and metrology in less developed countries. The UN Industrial Development Organization (UNIDO) has a long history of working with ISO to develop standardization capacity in developing countries.

Standards Development

Standards are developed by a wide variety of organizations and institutions in addition to the formal organizations already mentioned and this is particularly true in the ICT sector. The Institute of Electrical and Electronic Engineers (IEEE) is a professional membership society that develops standards through the IEEE Standards Association (IEEE SA).

Standards Development Organizations like IEEE have strict rules that must be followed in developing standards. IEEE SA's principles of standards development are: consensus, due process, openness, right to appeal and balance. All individuals and entities participating in IEEE SA standards development activities must abide by the IEEE Code of Ethics.

Organizations developing international standards must comply with The World Trade Organization (WTO) Committee on Technical Barriers to Trade (TBT) principles for standards development. The WTO principles are: transparency; openness; impartiality and consensus; effectiveness and relevance; coherence, and development dimension. There are international standards and International Standards with the term International Standard reserved for ISO and IEC standards.

ISO/IEC Directives, Part 2 states the general principles by which ISO and IEC documents are drafted. ISO and IEC committees are made up of experts on the standard topic. One of the advantages for governments of using ISO and IEC standards and standards in general is that the committees that develop standards are made up of content experts that represent the global community of practice relevant to the standard being developed.

The International Telecommunications Union (ITU) is the UN specialized agency for information and communications technologies (ICT). ITU has the lead role for the UN in making ICT work for the SDGs. ITU does its work through study groups of experts that develop technical standards or "Recommendations." The Recommendations are made freely available for industry and government to implement and operationalize.

Recommendations, reports and other publications can be downloaded from the ITU website.

Conformity Assessment

Conformity assessment is a process used to demonstrate that a product, service or system meets a standard. The main forms of conformity assessment are testing, certification, and inspection. Approaches to conformity assessment include: first party, second party, third party, regulation, and various combinations.

There is a process of mutual recognition so that certifications can be recognized globally. The International Accreditation Forum (IAF) recognizes accreditation bodies that meet specified ISO standards in accrediting certification bodies. IAF recognition is based on peer evaluation. The International Laboratory Accreditation Cooperation (ILAC) performs a similar function for laboratory (testing) and inspection accreditation. The goal is conformity assessments accepted globally.

Certification is typically voluntary but it can also be mandatory. For example, a government agency might require that a product or process be certified to meet a specific standard. Mandatory certification is common where public health and safety are involved. In some cases, governments or government agencies do the certification.

Standards in Governance

The focus of this chapter is technical standards but as already suggested, ISO produces a wide variety of management standards including *ISO 37001 - Anti-Bribery Management Systems*. This is a standard that connects directly to SDG 16 Governance - target 16.5 - Substantially reduce corruption and bribery in all their forms. *ISO 26000 - Social Responsibility* was initially developed to apply only to businesses but in its final form it applies to all organizations and its overall goal is sustainable development.

Standards are often incorporated into building codes by reference. Building codes are an important tool for improving building sustainability and resilience as well as ensuring the protection of public health and safety. The advantage to governments in using voluntary consensus standards is they take advantage of the technical expertise in standards committees - expertise that may not be available locally.

Buildings are large users of energy and large carbon emitters. Energy use and carbon emissions can be addressed in a building code or in a separate energy code. One of the ways that governments can directly impact the SDGs is by enforcing strict energy and carbon standards in construction and operation of facilities.

Another way that the government at all levels drives sustainable development is through their procurement practices. For the UN, sustainable procurement means taking societal and environmental factors into consideration along with finances. Common sustainability

goals of procurement are to reduce material use and carbon emissions. Some ways this can be done are by using sustainable vendors and by requiring vendors to meet minimum standards for recycled content and for decarbonization in goods and services. ISO has a guidance standard for sustainable procurement.

Artificial Intelligence Standards

One of the technologies that AI is expected to play an enabling role in is autonomous vehicles (AV) as part of intelligent transportation systems (ITS). ITS have been in development for more than 30 years and standardization has been and continues to be one of the challenges. ISO/ TC 204 Intelligent Transportation Systems was established in 1992 with a focus on ITS systems and infrastructure. ISO/TC 204 recently added an Ad Hoc Group on big data and artificial intelligence.

The Coursera course Smart Cities - Management of Smart Infrastructure includes a module on the digitalization of urban transportation systems. One of the takeaways from this course is that standards are critical to the digitalization of urban systems - smart cities - even before the addition of artificial intelligence.

In a recent paper, Cihon provided an overview of some of the issues around AI standardization. Cihon's focus is on ISO and IEEE but the paper also provides an introduction to standardization issues specific to AI. A major concern is that formal standards processes may or may not be giving enough attention to non-technical social issues such as privacy.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (the IEEE Global Initiative) is attempting to address this with its *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition (EAD1e)*. A key message in EAD1e is that ethical principles must be turned into practice.

The International Telecommunications Union (ITU) is playing an increasingly important role in the evolving world of IoT and AI. For digitally enabled systems to work, systems and system components must be able to communicate. The mission for ITU's AI for Good Global Summit series is "Leveraging AI to Accelerate Progress toward the SDGs."

AI and Machine Learning are gaining a larger share of the ITU standardization work programme in fields such as network orchestration and management, multimedia coding, service quality assessment, operational aspects of service provision and telecom management, cable networks, digital health, environmental efficiency, and autonomous driving.

Open platforms advancing various aspects of AI and Machine Learning

ITU have a few focus groups advancing different aspects of AI and Machine Learning. 1) Focus Group on 'AI for Health', driven in close collaboration by ITU and WHO, is working towards the establishment of a framework and associated processes for the performance benchmarking of AI for Health solutions. 2) Focus Group on 'AI for Autonomous and Assisted Driving' is working towards the establishment of international standards to monitor and assess the behavioral performance of the AI 'drivers' in control of automated vehicles. 3) Focus Group on 'Environmental Efficiency for AI and other Emerging Technologies' will benchmark best practices and describe pathways towards a standardized framework to assess environmental aspects of the adoption of emerging technologies. 4) Focus Group on 'Autonomous Networks', established in December 2020, will lead an exploratory 'pre-standardization' study to determine how ITU standards will support the realization of autonomous networks and their evolution in years to come. 5) Focus Group on 'AI for natural disaster management', established in December 2020, will support global efforts to improve our understanding and modelling of natural disasters in the interests of effective preparation and response. Multistakeholder initiative supported in close collaboration by ITU, the World Meteorological Organization (WMO), and the UN Environment Programme (UNEP).

The Global Initiative on 'AI and Data Commons' aims to support AI for Good projects in achieving global scale. The Initiative will offer assemblies of resources to launch new AI projects aligned with the SDGs, and scale them up fast.

The Global AI Summit, organized by the Saudi Data & AI Authority (SDAIA) brings together stakeholders from the public sector, the academia and the private sector to shape the future of Artificial Intelligence (AI). The summit was a unique global opportunity to explore what the new global era means to AI, how AI's transformational potential can be best used to create a better future for all, and what this means for policy and decision makers interested in leveraging this potential, for the good of humanity.

ISO/IEC JTC 1/SC 42 is the international standards committee responsible for standardization in the area of Artificial Intelligence (AI). It is setup as a joint committee between ISO and IEC, the international standards development organizations (SDOs). As the focal point of standardization on AI within ISO and IEC, SC 42's program work looks at the entire AI ecosystem. Additionally, SC 42 is scoped to provide guidance to ISO and IEC committees developing Artificial Intelligence applications. Its current program work includes standardization in the areas of foundational AI standards, Big Data, AI trustworthiness, use cases, applications, governance implications of AI, computational approaches of AI, ethical and societal concerns. The SC 42 web site includes a list of standards under development. This list is not a complete list but does cover IEEE, ISO, IEC, and ITU-T as of May 2019. For more on NIST's AI standards activities see the NIST AI website.

AI Standards Strategies

Only a few countries have standards strategies and even fewer have AI Standards strategies. Standards are included in some of the AI strategies discussed in Chapter 4 of this Guide. China addresses standards in its *Guidance on New Generation AI Development plan* and released *Guidelines for Construction of National New Generation Artificial Intelligence Standard System* to define the top-down design of China AI standard system. In *New Generation AI Governance Principles – Developing Responsible AI*, China calls for all countries to abide by ethics and standards and for consensus standards.

The European Union's "Policy and investment recommendations for trustworthy Artificial Intelligence" calls for the creation of recognized standards and fostering the development of standards for the interoperability of public applications and data sources. Also mentioned in the report of the High-Level Group is the need to consider a range of certification mechanisms for AI systems and the need for a clear standardization strategy to ensure trustworthy AI. In the European Commission's Coordinated Plan on Artificial Intelligence, the need for common standards is noted. In the statement on artificial intelligence robotics and autonomous systems, attention is drawn to the need for ethical guidelines that could serve as the basis for global standards. Germany is partnering with Deutsches Institute fur Normung (DIN) on its AI strategy.

The UK Industrial Strategy for AI mentions mobility, procurement, and security standards, and notes that the UK actively participates in international standards development especially in areas such as artificial intelligence and data protection. In a policy paper on AI, the UK government calls for industry to work for technical standards that support interoperability of AI systems and to work with the government to accomplish this. Japan sees international standardization as an important part of its AI strategy. India's concern is for data protection and India supports the adoption of international standards.

In 2019, NIST developed the U.S. AI standards strategy in response to the 2019 Executive Order for a National AI strategy. The NIST plan includes background specific to AI standardization. The plan notes that there are a number of cross-sector (horizontal) and sector-specific (vertical) AI technical standards already available, and many are under development. There is much less available in non-technical areas such as trustworthiness. AI standardization should make the maximum use of existing standards. Systems using AI technologies are generally systems of systems e.g. smart sustainable cities and AI standards should take this into account. Both systems AI standards and application specific AI standards are needed.

The NIST plan notes that for application specific AI standards, the U.S. Department of Transportation (DOT) and the U.S. Food and Drug Administration (FDA) are ahead of other U.S. agencies and departments in looking at AI standards.

The NIST plan calls for standardization projects that lead to "... globally relevant and non-discriminatory standards, where standards avoid becoming non-tariff trade barriers or locking in particular technologies or products." To not stifle innovation, standards should

have maximum flexibility, be platform neutral and be performance-based rather than prescriptive.

Standards Australia's *Artificial Intelligence Standards Roadmap: Making Australia's Voice Heard* includes recommendations for AI ranging from safety and trust for citizens and consumers to opportunities to enhance export opportunities and calls for Australians to play their part to shape the development of standards for AI internationally.²¹

Key References

Following is a summary of some recent standards publications available online.

1. An Introduction to Standardization: A practical guide for small businesses

The first section of the guide outlines the economic benefits of standardization. The voluntary nature of standards supports self-regulation by industry and reduces the legislative burden on the government. Standards provide internationally recognized solutions for safety, health and environmental protection. The use of internationally recognized standards facilitates trade.

Although standards can be a barrier to innovation, they can also be an enabler in moving innovation into the marketplace particularly where prospective users look for compatibility and interoperability. An example is given of charging systems for electric vehicles.

How standards are developed is the subject of the second section. Standards developed by the German Institute for Standardization (DIN) are designated DIN. DIN standards can result from work at the national, regional, or international level and are designated accordingly. Examples are given for the different DIN standards designations followed by a list of useful terms.

[Download here](#)

2. AI Global: Responsible AI Community Portal

The Responsible AI Community Portal is an evolving repository of reports, standards, models, government policies, datasets, and open-source software to inform and support responsible AI development. These resources are collected and classified by the NGO AI Global. Resources are often scattered on different platforms and can be difficult to find. The purpose of the Community Portal is to allow quick and easy access to 300+ resources that bring awareness to ethical AI issues and how to mitigate harm of technology.

²¹ Citations for this chapter:

UNECE. (2018, September 26). Standards for the Sustainable Development Goals, Geneva CIG. Retrieved from <https://www.unece.org/sdgs-isoweek2018.html>

Artificial intelligence. (2020, September 02). National Institute of Standards and Technology. Retrieved from <https://www.nist.gov/artificial-intelligence>

By reducing the time, it takes to search for trusted resources, the goal is to improve access and empower progress in this rapidly evolving field. Users can browse through selected list of resources classified by the AI Global team and can search terms or apply a variety of filters, which include pathways users can select depending on their role. Anyone is able to search terms and access the resources but to upload a resource, users need to create an account. Users can then share the resource, which will first be reviewed by AI Global admins before being displayed on the Community Portal. AI Global's mission is to mitigate harm and unintended consequences of technology by building practical tools to support the responsible development of AI systems for individuals and teams. The platform was built to support the community working on the development of AI, including developers, product designers, researchers, business owners, policymakers, and risk managers. The objective is to engage people from different perspectives and roles - it is only through collective effort and knowledge that progress can be made. By bringing resources catered to people from a variety of backgrounds, AI Global intends to help people to work more effectively in developing Responsible AI.

[Download here](#)

3. ABCs of Conformity Assessment

This publication provides an overview of the conformity assessment. It describes conformity assessment terminology and concepts, identifies some of the interrelationships among conformity assessment activities, and discusses possible impacts on trade. Conformity assessment is defined in ISO/IEC 170001 as the "demonstration that specified requirements relating to a product, process, system, person or body are fulfilled".²²

Conformity assessment procedures provide a means of assuring that products, services, or systems produced or operated have the required characteristics. Conformity assessment includes testing and inspection, as well as certification of products, management systems, and personnel. It also includes accreditation of organizations performing conformity assessment activities.

Within ISO, standards related to conformity assessment are developed and published by the ISO Committee on Conformity Assessment (CASCO). The conformity assessment standards are commonly known as the CASCO toolbox.²³ These conformity assessment standards are developed and published jointly by ISO and IEC. The CASCO toolbox is recognized and used globally

Accreditation provides confidence, through an independent evaluation of conformity assessment bodies against standards to carry out specific activities, that conformity assessment organizations meet requirements and operate with independence,

²² ISO/IEC 17000 "Conformity Assessment – vocabulary and general principles" provides terms and definitions applicable to conformity assessment. <https://www.iso.org/obp/ui/#iso:std:iso-iec:17000:ed-1:v1:en> Accessed May 23, 2020

²³ ISO CASCO Toolbox https://www.iso.org/sites/cascoregulators/02_casco_toolbox.html Accessed May 23, 2020

impartiality, and competence. There are accreditation programs for testing laboratories, and inspection bodies, as well as certifiers.

Conformity assessment procedures are important for global trade. The World Trade Organization (WTO) Technical Barriers to Trade (TBT) Agreement contains obligations regarding conformity assessment procedures and their use in international trade.

Mutual recognition agreements and mutual recognition arrangements are used to facilitate the acceptance of conformity assessment results between two or more parties. Accreditation bodies from around the world have formed international and regional cooperation and established “multilateral agreements or arrangements” to recognize and accept the results of conformity assessment.

[Download here](#)

4. An Artificial Intelligence Standards Roadmap: Making Australia’s Voice Heard

This Roadmap, developed by Standards Australia (SA) the national standards body, is a result of consultation with a broad cross-section of stakeholders. It is intended to provide the framework for Australians to intervene and shape the development of standards for AI internationally.

This Roadmap builds on a growing body of work globally on approaches to managing the impact of AI including how it might impact and enable the United Nations Sustainable Development Goals. It builds on work in the United States, the United Kingdom, Singapore, the New Zealand AI Forum and the ongoing work of the Australian Human Rights Commission on the human rights impacts of new technologies including AI.

An overview of the role standards can play in managing the development and adoption of AI, using examples from the digital economy is provided. International AI standards work underway within Standards Development Organizations (SDOs), specifically ISO/IEC JTC 1/SC 42 Artificial Intelligence, and other multilateral and commitments the Australian Government has made are summarized. Australia is committed to the development of consensus-driven Standards on AI, through the OECD Principles on AI. SA sees the OECD's call for governments to promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI as an encouragement to participate in ISO and IEC standardization work.

Ideas and feedback Australian stakeholders provided on AI Standards are summarized. Examples include identifying specific opportunities for Australia to play a leadership role in international SDOs, the need to focus on specific issues, such as privacy and inclusion and fairness, and to adopt a balanced approach in policy and regulation. Privacy was a key theme in a number of submissions, from businesses, consumers and government agencies. Standards can play a strong role in promoting inclusive design and the use of AI consistent with laws or good practices. A range of submitters and workshop participants raised the role standards for AI could play in preventing and addressing discrimination, improving the accuracy of services, ensuring inclusion, safeguarding democracy, and building trust. A number of stakeholders proposed certification models for AI to shape responsible AI. Ideas and feedback are captured in seven specific recommendations.

The report concludes noting the important opportunities AI presents for Australians and calling for the private sector, civil society and the Government to work together on common goals for AI.

[Download here](#)

5. Ethically Aligned Design First Edition Overview

Ethically Aligned Design is mainly about scientific analysis and resources, high-level principles, and actionable recommendations for AI. It also offers specific guidance for standards, certification, regulation, and legislation for design, manufacture, and use of automated and intelligent systems (A/IS) that provably aligns with and improves holistic societal well-being.

There are a number of standards projects related to the IEEE Global Initiative that produced *Ethically Aligned Design*. The IEEE P7000™ series of standardization projects explicitly focus on societal and ethical issues associated with AI. This is a first since most IEEE standards projects focus on technical issues like efficiency and interoperability.

An A/IS Ethics Glossary has been developed and is available for download on the IEEE website. The glossary is intended for a broad audience of stakeholders including engineers, policymakers, philosophers, AI researchers, and standards developers.

As mentioned earlier, the IEEE P7000™ series of standards projects under development are different from those normally developed by IEEE SA with their focus on issues at the intersection of technological and ethical considerations. Examples of projects are given. A current list of standards projects and their status and how to get more information or join a standards project working group under development can be accessed on the Initiative website.²⁴

[Download here](#)

6. Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0 (AV 4.0)

Realizing the full potential of AVs will require collaboration and information sharing among stakeholders from industry, State, local, tribal, and territorial governments, academia, not-for-profit organizations, standards development organizations (SDO), and the Federal Government.

The U.S. government has established ten principles in three core areas for Automated Vehicles (AV). The three areas are: protect users and communities; promote efficient markets; and facilitate coordinated efforts. Under coordination is the principle to promote consistent standards and policies. A number of the principles have standards of implications or aspects.

For standards, the U.S. Government will prioritize participation in and advocate internationally for voluntary consensus standards and evidence-based and data-driven regulations. The U.S. Government will engage State, local, tribal and territorial authorities as well as industry to promote the development and implementation of voluntary consensus standards, advance policies supporting the integration of AVs throughout the

²⁴ IEEE Ethics in Action <https://ethicsinaction.ieee.org/>

transportation system and seek harmonized technical standards and regulatory policies with international partners.

The U.S. Government will promote voluntary consensus standards as a mechanism to encourage increased investment and bring cost-effective innovation to the market more quickly.

There are three appendices. Appendix A is a list of the U.S. Government resources related to AV. The second is a list of US. Government AV contacts and the third is a list of the members of the Automated Vehicle Fast Track Action committee.

[Download here](#)

7. Fast Forward: National Standards Bodies in Developing Countries

This publication is a joint effort of ISO and the UN Industrial Development Organizations (UNIDO) to support metrology, accreditation and standardization - a quality infrastructure system - in developing countries. It covers the main principles of standardization at national, regional and international levels and illustrates the elements of quality infrastructure management at a national level. The publication is intended to support the establishment and development of the National Standards Bodies (NSB).

Part one outlines in some detail, the three pillars of a quality infrastructure system: metrology, standardization, and conformity assessment and accreditation. It is noted that having a fully functional quality infrastructure system may be beyond the resources of some developing countries. Some different ways that developing countries can access this capacity are outlined.

Part two outlines the role that the WTO plays in standardization. Part three; four, five, and six cover standards, standards bodies, national standards bodies, and standards development. Part seven outlines different ways that NSBs serve stakeholders. The main function of a NSB is monitoring international standards activity that relates to the national economy and providing accurate and timely information. This will typically involve the NSB selling international standards and this can be a source of revenue and support other NSB services.

Part eight deals with regional and international relations and how these can support the work of NSBs. Many resources are available to assist developing countries in developing and maintaining an NSB and the quality infrastructure system necessary to support increased trade and sustainable development.

[Download here](#)

8. OCEANIS: The Open Community for Ethics in Autonomous and Intelligent Systems

OCEANIS provides a high-level global forum for discussion, debate and collaboration for organizations interested in the development and use of standards to further the development of autonomous and intelligent systems. The Global AI Standards Repository is the world's first centralized, transparent notification system that captures AI and Autonomous and Intelligent Systems standards and standards in progress.

[Download here](#)

9. One AI: "What Tools for Trustworthy AI? A preliminary framework"

This report provides a preliminary framework to categorise and compare tools for implementing trustworthy AI. The framework is the outcome of analysis by the OECD Network of Experts on AI (ONE AI) on existing technical approaches to implementing trustworthy AI (e.g. technical standards, toolkits, etc.); process-related approaches (e.g. guidelines, risk management tools, process standards, etc.); and educational and capacity building approaches (e.g. change management processes, educational material, etc.). The framework will serve as the basis for an open access and interactive database available on [OECD.AI](https://oecd.ai).

Forthcoming, May 2021

10. Standards and Standardization: A practical guide for researchers

This guide was prepared for participants in European research projects to inform them about opportunities to use standardization for disseminating research results. The first part of the guide outlines standards development. The second part provides guidance for researchers on the opportunities, procedures and value of standardization.

It is pointed out that simply because information about standards does not exist in the relevant academic literature it still may exist since there is a significant gap between academe and practice. However, there are many resources available on the Internet for identifying relevant standards resources.

There is a brief discussion of research results that might be appropriate for standardization recognizing that these are very sector specific and a discussion of standards versus patents. Guidance is given on what type of standard project and what organizations might be appropriate. Annex A of the guide is a template outlining the stages in taking a research output to a published standard.

Part three of the guide provides details of the standardization process for researchers.

In part four, the standards review and approval process are outlined. A critical part of the standardization process is providing for comments and resolving comments. Part five gives some examples of successful standardization projects resulting from European Framework Projects. There are six annexes with details on how to take a standards project from start to finish.

[Download here](#)

11. Technical Report Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development

Standards, particularly those developed by existing international standards bodies, can support the global governance of AI development. There is a well-developed institutional structure for international standards development and international standards bodies have a track record of developing a wide range of standards important for AI including cybersecurity, environmental sustainability, and safety including areas such as autonomous vehicles and nuclear energy. Standards bodies have the institutional capacity to achieve expert consensus and promulgate standards globally. These standards can then be used voluntarily or in regulation.

Ongoing standards work has focused primarily on standards to improve market efficiency and to a lesser extent on ethical concerns. There remain risk standards could fail to address important policy objectives and that AI research organizations that could contribute may not participate in standardization efforts.

Standards cannot support all AI policy goals, but they are an important part of effective global solutions. Standards influence the development and deployment of AI systems through product specifications for example, explainability, robustness, and fail-safe design. They can also affect the larger context in which AI is researched, developed, and deployed through process specifications. The creation, dissemination, and enforcement of international standards can help build trust among participating researchers, labs, and states.

Standards serve to disseminate best practices globally. Existing international treaties, national mandates, government procurement requirements, market incentives, and global harmonization pressures can all contribute to the spread of standards once they are established.

Standards do have limits, however: existing market forces are insufficient to incentivize the adoption of standards that govern fundamental research and other transaction-distant systems and practices. Concerted efforts among the AI community and external stakeholders will be needed to achieve such standards in practice.

Ultimately, standards are a tool for global governance, but one that requires institutional entrepreneurs to actively use standards in order to promote beneficial outcomes. Key governments, including China and the U.S., have stated priorities for developing international AI standards. Standardization efforts are only beginning, and could become increasingly contentious over time, as has been witnessed in telecommunications. Engagement sooner rather than later can establish beneficial and internationally legitimate ground rules to reduce risks in international and market competition for the development of increasingly capable AI systems. In light of the strengths and limitations of standards, this paper offers a series of recommendations.

[Download here](#)

12.Using and referencing ISO and IEC standards to support public policy

This guide was prepared by IEC and ISO as a way of making their International Standards more visible to public policymakers. ISO and IEC standards are "voluntary" but they can be valuable public policy tools. The primary audience for this guide is national decision-makers and their national member standards bodies.

The introduction begins with a brief description of what an IEC or ISO International Standard is and why they should be important for policymakers. One reason is that they can be powerful tools for good governance. Secondly, it is important because of the need to align policy making with good standardization practice. This guide focuses on ISO and IEC standards but notes that there are other international standards; for example, IEEE.

The standards development process is outlined and it is noted that ISO provides guidance for committees developing standards related to public policy issues.²⁵ An advantage of ISO and IEC standards is that they can be adopted as national standards (with or without modification) after completion of a national public enquiry process.

The guide goes into some detail on how policymakers can use standards in legislative (e.g. technical regulations supporting laws) and non-legislative (e.g. public procurement) actions. Some specific examples are given on the general use of standards by policymakers and for procurement. The point is made that because governments are major procurers of goods and services that this is a good way to drive policy implementation.

Section four outlines how to reference ISO and IEC standards once the decision to use a standard has been made. Standards maintenance and conformity assessment are the subjects of section five. Policymakers can specify how and who does the conformity assessment to meet a specific requirement. Regulators may do the conformity assessment or they could require third-party assessment; examples are given.

National policies for the use of standards to support public policy are outlined in section seven. Some countries allow referencing of ISO or IEC standards without adoption whereas others require that the standards must first be adopted as national standards. Examples are given for how standards are designated that have been regionally or nationally adopted. Examples of national policies for Brazil, Canada, China, the European Union, Japan, Mexico, South Africa, and the United States are given; more examples are available in ISO/IEC Guide 21-1.²⁶ Section seven is for standards supporting public policy in different sectors. A list is given of the sectors but readers are referred to a website for the examples.^{27 28}

[Download here](#)

13. UNECE Standards for the SDGs

This publication provides an overview of how international standards are being used and can be used by policymakers to support sustainability and the achievement of the SDGs. Case studies are presented to illustrate the use of standards for: SDG 6 Clean Water and Sanitation; SDG 7 Affordable and Clean Energy; SDG 11 Sustainable Cities and Communities; and SDG 13 Climate Action.

Voluntary national and international standards support the achievement of the 2030 Agenda in different ways. Case histories illustrate how standards support specific goals and targets. Standards are used in design and manufacturing of products and can also be used in the regulations of products. For example, products that support the achievement of SDG 7 by improving energy efficiency.

²⁵ ISO IEC Annex SO of the Consolidated ISO Supplement to the ISO/IEC Directives Part 1

https://www.iso.org/sites/directives/current/consolidated/index.xhtml#_idTextAnchor618 Accessed May 19, 2020

²⁶ ISO/IEC Guide 21-1, Regional or national adoption of International Standards and other International Deliverables – Part 1: Adoption of International Standards <https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:21:-1:ed-1:v1:en> Accessed May 19, 2020.

²⁷ ISO Examples by Sector https://www.iso.org/sites/policy/sectorial_examples.html Accessed May 19, 2020

²⁸ IEC Examples by Product Sector <https://www.iec.ch/perspectives/government/sectors/> Accessed May 19, 2020

Smart cities and intelligent transportation systems are expected to play an important role in achieving SDG 11. These systems of systems rely on a high level of interoperability between the infrastructure and the services. AI standards will build on and enhance these systems.

[Download here](#)

14. U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools

The EO directed the Secretary of Commerce, through the National Institute of Standards and Technology (NIST), to issue “a plan for Federal engagement in the development of technical standards and related tools in support of reliable, robust, and trustworthy systems that use AI technologies.” NIST supports a wide variety of standards activities for the Federal government and NIST staff activity participate in standards development organizations including IEEE, ISO, and IEC.

There are a number of cross-sector (horizontal) and sector-specific (vertical) AI standards available now and many others are in various stages of development. Horizontal standards can be used across many applications and industries. Standards developed for specific application areas such as healthcare or transportation are vertical standards. Some areas, such as communications, have well-established and regularly maintained standards in widespread use. Systems using AI technologies are generally systems of systems, and AI standards should take this into account. AI standards encompass those specific to AI applications as well as standards for components of an AI-driven system—and both types of standards are needed. Technical standards can provide developers clear guidelines for the design of AI systems to ensure that they can be easily integrated with other technologies, utilize best practices for cybersecurity and safety, and adhere to a variety of technical specifications that maximize their utility.

The NIST plan includes detailed background on technical standards and how they are developed and a section on what AI standards are needed. The plan identified nine areas of focus for AI standards including trustworthiness. Trustworthiness standards include guidance and requirements for accuracy, explainability, resiliency, safety, reliability, objectivity, and security.

It is widely agreed that societal and ethical issues, governance, and privacy are issues that must be addressed in AI standards but it is not clear yet how this is to be done. Also, some aspects of AI development do not lend themselves to standardization such as mathematical and statistical theory.

The plan also calls for the development of tools to support the development of AI technologies including testing methodologies to validate and evaluate AI technologies and AI testbeds. The plan recommends that the Federal government actively support AI standards development activities to help speed AI technology development and maximum use should be made of existing standards.

Specific recommendations for the Federal government are offered on coordination, research, partnerships and international engagement.

[Download here](#)

Chapter 4: National, regional and international AI strategies

Introduction

How countries react to and support the development of AI will have far-reaching impacts, including on global competitiveness. Therefore, governments have started to develop national strategies on AI, in order to demonstrate leadership in guiding national AI development. This chapter provides an overview of the global landscape of different national and international AI strategies and the key features of these documents. It is notable that AI strategies seek to not only strengthen the development of AI, but also to drive the development of related sectors within the overall economy, as well as to standardize the application of AI through regulation or frameworks to address ethical concerns.

In an increasingly digital and data-driven world, AI may very well bring on a new era. Undoubtedly, AI should be a force for good, enabling groundbreaking insights and applications. Yet AI poses significant risks. As we face the exciting prospects that AI can usher in, we must consider all of AI's implications for the future, both positive and negative. In order to better understand how AI may shape the future, especially in accelerating the achievement of the SDGs, policy makers will need to look at how the field has evolved and consider where AI stands today and all implications for the future, both positive and negative.

The following observations are identified by and for policy makers:

Limited accessibility of AI research

A lack of accessibility to AI research puts limit on accountability and the degree of reproducibility of AI. Exceptions include the PostEra COVID Moonshot²⁹, an open sourced and crowdsourced collaboration using machine learning to accelerate drug design and testing. In fact, the field of biology has recently taken up AI in many sectors, including medical imaging, genetics, proteomics, chemistry, and as aforementioned, drug discovery.

Natural language processing, which enables machines to analyze, understand and manipulate language, has been a major focus of AI research, predominantly taking place in large companies with huge models and training costs. While AI has traditionally had an open ethos, the industrialization of AI serves to centralize AI talent and diminish accessibility in order to preserve intellectual property.

Lack of transparency

Related to the limited accessibility of AI research are questions related to transparency. As observed above, many AI-developed technologies rely on machine learning algorithmic

²⁹ <https://covid.postera.ai/covid>

processes that result in a “black box”. Such opacity hinders detection of bias and discrimination, ultimately increasing potential to harm. Areas for policy debate include algorithmic transparency, transparency in data used and held, transparency related to content moderation decisions, and transparency in review. Transparency requirements could be made open for the public or qualified for specified authorities/bodies.

Corporate-driven brain drain

Currently, top-tier AI research has been dominated by several countries, whose universities and corporations have been leading in acceptances of major academic conference papers. This lead is driven by recruitment of international talent.

Private corporations are recruiting more and more tenured and tenure-track professors. Simultaneously, there have been several new institutions of higher education dedicated to AI that have been formed around the world.

Pandemic further - accelerates AI ethical concerns

During COVID-19 crisis, millions have died, tens of millions pushed into poverty and hunger, and inequalities exacerbated along many dimensions. The pandemic has demonstrated the promise and power of technology to alleviate its impacts, while simultaneously heightening the disquieting issues of lack of access and invasion of privacy.

The global health crisis also has emphasized how vital it is for countries and societies to advance their technology capabilities, and related policy capacities, if they are to not be left behind. AI can enable faster progress on many goals and targets through innovative new solutions, more efficient resource use, and better decision making using big-data analytics.

On the other hand, AI can perpetuate biases and inequalities; or even be used intentionally to violate human rights and develop applications that harm individuals and societies. Like many powerful technologies, AI, too, must be guided by human values, human solidarity, human intelligence and respect for human rights.

Key References

Below is a summary of some relevant publications related to AI strategies and policy recommendations, which are available online.

National Strategies

1. Argentina: Artificial intelligence and data science, potential and challenges for the management of health crises

The cycle of meetings "Artificial intelligence and data science, potential and challenges for the management of health crises", gathers specialists in the area discussed and

reflects on the role of information and communication technologies, artificial intelligence, and data science in the context of the pandemic.

The cycle of meetings was organized by the Ministry of Science, Technology and Innovation, the Ministry of Health, the National Agency for the Promotion of Research, Technological Development and Innovation (I + D + I Agency), and the Sadosky Foundation. The theme of the first meeting was "ICT tools applied to smart health management".

During the COVID-19 pandemic it became evident that there is a need to create new strategies to govern risks, generating trust and predictability. The world has warned that information and communication technologies (ICT), artificial intelligence (AI) and data science (Big Data, predictive analytics and data mining), constitute a real opportunity for the development of innovations and technologies suitable for crisis management.

In Argentina, a variety of applications, systems, technologies and innovations associated with artificial intelligence and data science have been developed by both public and private actors to contain and mitigate the spread of COVID-19 and, based on these answers and new demands that arise, have unleashed very pertinent questions. The purpose of this cycle of meetings is to discuss where the country is today in this matter, what capacities it has, what the challenges are, and what ethical debates exist.

[Download here](#)

2. Canada: Directive on Automated Decision-Making

Canada is one of the countries, which has been active and vocal about responsible AI and the need for ethics to guide the development of AI. To this end, their government has committed to guiding principles to ensure the effective and ethical use of AI. The principles incorporate ethical values but also provide concrete actionable points for the government. As part of this commitment, the Government of Canada introduced the new Directive on Automated Decision-Making, which took effect on 1 April 2019.

The Directive sets out minimum requirements for federal government departments that wish to use an Automated Decision System, and applies to any system, tool or statistical model in production used to recommend or make an administration decision about a client. The Directive outlines the following requirement: (1) Algorithmic Impact Assessment, (2) Transparency, (3) Quality Assurance, (4) Recourse, and (5) Reporting.

[Download here](#)

3. China: New Generation Artificial Intelligence Development Plan

In July 2017, the Chinese Ministry of Industry and Information Technology presented a national strategy for Artificial Intelligence, *New Generation Artificial Intelligence Development Plan*. The Chinese government regards Artificial Intelligence as a key industry for the future and sets three-step strategic objectives:

(1) by 2020, the overall technology and application of AI will be in step with globally advanced levels, the AI industry will have become a new important economic growth point, and AI technology applications will have become a new way to improve people's livelihoods.

(2) by 2025, China will achieve major breakthroughs in basic theories for AI, and AI becomes the main driving force for China's industrial upgrading and economic transformation.

(3) by 2030, China's AI theories, technologies, and applications should achieve world-leading levels, making China the world's primary AI innovation center, laying an important foundation for becoming a leading innovation-style nation and an economic power.

Six focus tasks are emphasized in this national strategy: (1) Build open and coordinated AI science and technology innovation systems. (2) Fostering a high-end, highly efficient smart economy. (3) Construct a safe and convenient intelligent society. (4) Strengthen military-civilian integration in the AI domain. (5) Build a safe and efficient intelligent infrastructure system. (6) Plan a new generation of AI major science and technology projects

There are also chapters regarding resource allocation, guarantee measures and organization and implementation to support this national AI strategy.

[Download here](#)

4. Chile: Artificial Intelligence Strategy

The objective of the Chile's national AI strategy published in April 2021, is to empower the country in use and development of AI systems and fostering debates about its ethical dilemmas and its regulatory, social and economic consequences. It also proposes building a path on how we will change as a society with this technology in the next few years. During 2020, a committee of experts, together with other ministries, worked in coordination with the future team of the Ministry of Science, Technology, Knowledge and Innovation in order to prepare the national strategy. More than 1,700 people participated in self-convened tables and workshops deployed throughout the country. In addition, 15 webinars were generated to bring AI topics closer to the community in general, reaching an audience of 5,400 people. After the participation process described above, this document was taken to public consultation late 2020. The cross-cutting principles of this policy are based on a human-centered AI which is safe, inclusive, globalized, at the service of society and generates sustainable development. The axes covered by this policy are: (i) enabling factors (talent development, technological infrastructure, and access to quality data); (ii) development and use of AI and (iii) ethical, legal, regulatory aspects and socio-economic impact. The roadmap plans more than 170 initiatives organized into 60 priority actions.

[Download here.](#)



5. Egypt: National AI Strategy

As artificial intelligence will be the driving force for Egypt's Vision 2030, Egypt issued its National AI Strategy in 2019. The strategy—to be implemented within three to five years— identifies the country's plans to deepen the use of AI technologies and transform the economy. The strategy aims at unleashing the potential capabilities of Artificial Intelligence in different focus areas: Education, Research and Development, Healthcare, Transportation, Fourth Industrial Revolution, Precision Agriculture, Financial Services.

Egypt's AI Strategy consists of 4 pillars: (1) AI for Government, (2) AI for Development, (3) Capacity Building, and (4) International Relations. It also identifies 4 enablers: (1) Governance, (2) Data, (3) Ecosystem, and (4) Infrastructure.

To achieve this strategy, Egypt has established the National Council of Artificial Intelligence in accordance with the decision of the Prime Council Minister to follow and govern the strategy implementation through the coordination between the relevant authorities.

[Available here](#)

6. Germany: Strategy for Artificial Intelligence: AI made in Germany

In November 2018, the German Federal Government announced its national *Strategy for Artificial Intelligence*. The goal is to establish “AI made in Germany” as an international trademark for cutting-edge, secure AI applications aimed at serving the common good in line with Europe's core values. There are 3 political goals: (1) make Germany and Europe a leading center for AI and thus help safeguard Germany's competitiveness in the future. (2) achieve a responsible development and use of AI which serves the good of society. (3) integrate AI in society in ethical, legal, cultural and institutional terms in the context of a broad societal dialogue and active political measures.

The Federation allocated a total of €500 million to beef up the AI strategy for 2019 and the following years. Up to and including 2025, the Federation intends to provide around €3 billion for the implementation of the Strategy.

There are 12 fields of action: (1) Strengthening research in Germany and Europe in order to be drivers of innovation. (2) Innovation competitions and European innovation clusters. (3) Transfer to the economy, strengthen Mittelstand (4) Fostering the founding of new businesses and leading them to success (5) World of work and the labor market: shaping structural change (6) Strengthening vocational training and attracting skilled labor/experts (7) Use AI for tasks reserved for the state and administrative tasks (8) Making data available and facilitating its use (9) Adjusting the regulatory framework (10) Setting standards (11) National and international networking (12) Engaging in dialogue with society and continuing the development of the framework for policy action.

[Download here](#)

7. India: National Strategy for Artificial Intelligence

Recognising AI's potential to transform economies and the need for India to strategize its approach, Hon'ble Finance Minister, in his budget speech for 2018 – 2019, mandated NITI Aayog to establish the National Program on AI, with a view to guiding the research and development in new and emerging technologies. NITI Aayog after having round of consultations released a discuss the government in 4 overarching areas: (1) Research

and Application, (2) Reskilling and Training, (3) Accelerating Adoption of AI, and (4) Responsible AI Development.

[Download here](#)

8. Italy AI Strategy: Artificial Intelligence at The Service of The Citizen

This white paper is the result of the work of the Agency for Digital Italy with a task force of experts, aims to analyse the impact of Artificial Intelligence in our society and more specifically, in the public administration. Its goal is to facilitate the adoption of these technologies by the state, to improve services to citizens and businesses, thus giving a decisive boost to innovation, the smooth functioning of the economy and, more generally, progress in everyday life.

The Italian AI strategy is largely connected with the European and international vision on the subject. The Italian AI national strategy should pursue the following:

1. Increase public and private investments in AI and related technologies
2. Strengthen the research and innovation ecosystem in the field of AI.
3. Support the adoption of digital technologies based on AI.
4. Strengthen the educational offer at every level, to bring the IA to work force.
5. Exploit the potential of the data economy, real fuel for AI.
6. Consolidate the regulatory and ethical framework that governs the development of AI.
7. Promote awareness and trust in AI among citizens.
8. Relaunch the public administration and make public policies more efficient
9. Promote European and international cooperation for a responsible and inclusive AI.

The public-private partnership, the intra- and inter-institutional collaboration, the synergy with the European action and the monitoring and evaluation of the evidence generated will be the four cardinal points that will guide the Government in implementing the strategy. This strategy is part of the European Coordinated Plan for Artificial Intelligence and can be understood as a contribution to a synergic action between the countries of the European Union, with the hope that Europe will soon be able to compete with the United States and China.

The connection with the international forums for discussion and negotiation represents a foundation of the Italian strategy, largely inspired by the comparison in the EU and OECD (OECD, 2019: Artificial Intelligence in Society) , as well as by the cooperation within the G7 and the G20, and by the contribution of some Italian experts to European tables such as the “High-Level Expert Group on Artificial Intelligence” 13, the “AI Alliance” 14, and to the main laboratories on IA belonging to the research network of the CLAIRE confederation”.

[Download here](#)

9. Japan: Artificial Intelligence Technology Strategy



The *Artificial Intelligence Technology Strategy* was published by The Strategic Council for AI Technology in March 2017, which was founded in April 2016 to promote Artificial Intelligence in Japan. The goals and measures contained therein support the vision of a super-smart “Society 5.0”, as pursued by Prime Minister Shinzō Abe for several years with a package of policies (“Abenomics”).

This strategy divides the process of AI industrialization into three phases. Phase 1: Utilization and application of data-driven AI developed in various domains. Phase 2: Public use of AI and data developed across various domains. Phase 3: Ecosystem built by connecting multiplying domains.

The four priority areas are (1) productivity; (2) health, medical care and welfare; (3) mobility; (4) information security. The action plans are (1) promoting R&D led by three focal centers and based on Industry-Academia-Government Collaboration; (2) fostering of Human Resources; (3) strengthening environmental maintenance of data and tools owned by industry, academia, and government; (4) strengthening start-up support through open innovation and fostering human resources; (5) promoting understanding related to development of AI technology.

[Download here](#)

10. Japan: AI Strategy 2019 – AI for Everyone: People, Industries, Regions and Governments

Given that potential fields for introducing AI technology is so wide, competition in areas such as data collection and utilization in the field has just begun, and the decisive contest is yet to come, Japan government published this *AI strategy 2019* in June 2019, focusing on measures that Japanese government should immediately take concerted action on. The purpose of this Strategy is to specify the environment and measures conducive to effective future utilization of AI for the purposes of contributing to the solution of global issues through realization of Society 5.0 and overcoming the issues facing Japanese society. This Strategy sets out 4 strategic objectives:

(1) For Japan to develop a base of human resources, which, in proportion to population, leads the world in being aligned with the needs of the AI era, and to become a country that attracts human resources from around the world. In addition, to build a mechanism to achieve this object on a sustainable basis.

(2) For Japan to become a frontrunner in the application of AI to real-world industry and to achieve strengthened industrial competitiveness.

(3) For a series of technology systems to be established in Japan that will realize a "sustainable society that incorporates diversity", and to implement a mechanism to operate them.

(4) For Japan to take a leadership role in building international research, education, and social infrastructure networks in the AI field, and to accelerate AI-related R&D, human resource development, achievement of SDGs, etc.

To support these strategic objectives, this Strategy establishes an integrated policy package for AI that encompasses educational reform, research and development (R&D), social implementation, data-related infrastructure construction and AI era digital

government in order to contribute to the world, overcome challenges, and ultimately improve Japan's industrial competitiveness.

[Download here](#)

11. Mauritius: Artificial Intelligence Strategy

Mauritius became one of the first African countries to publish its 70-page AI national strategy in November 2018. The Mauritius Artificial Strategy follows the government's dedication towards making AI a cornerstone of the next development model by recognizing the potential of the technology to improve growth, productivity and the quality of life, and makes recommendations to achieve objectives set.

The strategy is based on the identification of priority projects that are AI-enabled in the various sectors of the economy, skill attraction/capacity building for efficient and effective adoption, incentives to catalyze implementation and adoption of new technologies for improved public services delivery to support AI implementation. In particular, focus is given to sectors such as manufacturing, healthcare, fintech, agriculture and ocean economy.

The strategy includes recommendations in the following areas: (1) Mauritius Artificial Intelligence Council, (2) Funding and incentives, (3) Skills development, (4) Projects, (5) Regulatory framework, ethics and data protection, and (6) Networks and clusters.

[Download here](#)

12: Mexico National AI Strategy: Multi-sectoral collaboration with public consultations for continuity

Mexico published its national AI strategy in 2018, the first Latin American country to do so. Its national agenda on AI, a follow up, was published in September 2020. Mexico's national AI strategy promotes a multi-sectoral, participatory approach to AI investment and development. Recommendations include the development of an AI subcommittee within the Mexican Intergovernmental Commission for Electronic Governance to promote multi-sectoral dialogue and approaches, to work with experts and citizens through the AI subcommittee, and to open up recommendations for public consultation. The AI subcommittee furthermore serves to ensure continuity of AI related efforts with future governments.



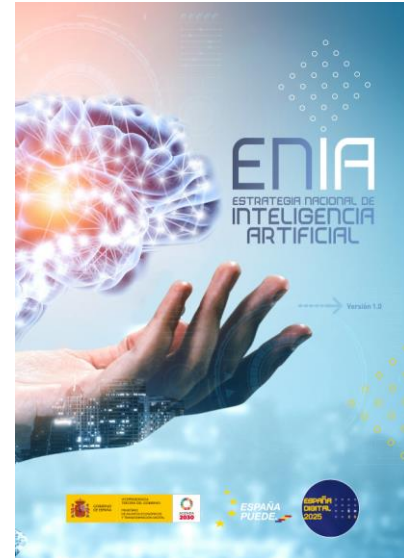
The national AI strategy looks to promote Mexico's international leadership in digital policy, with a special emphasis on the OECD and D7. Mexico's focus areas include data, digital Infrastructure and cybersecurity, ethics, research and development, governance, government and public services, skills, capacities and education and Mexicans abroad. Case studies of AI in Mexico demonstrate the use of emerging technologies to achieve social goals, such as increasing financial inclusion, combating corruption, improving public

health, and reducing crime. As such, Mexico has a strong focus on the social impacts of AI.

[Download here](#)

13.Spain: National strategy in Artificial Intelligence

Following the Spanish strategy of R&D in Artificial Intelligence, the Spanish government announced in December 2020 the AI National Strategy (ENIA), with objectives aligned with the 2030 plan and following the OECD recommendations. The design considers principles of robustness, safety, and unbiasedness, focusing on integrating economic, social, and environmental fields. The ENIA declares seven objectives covering excellence in R&D, the creation of qualified employments, the transformation of the productive matrix, trustworthy AI, human-centered AI, and inclusive and sustainable AI. Interestingly, ENIA projects Spanish as a mother language in technology development. Finally, an action plan presents six strategic aims: (i) to enhance R&D and innovation in AI; (ii) to promote the development of digital capabilities, enhance national talent and attract global talent in AI; (iii) to develop data platforms and technological infrastructures to support AI; (iv) to integrate AI into value chains for economic diversification; (v) to enhance the use of AI in the public sector and national strategic missions; (vi) to establish an ethical and regulatory framework that reinforces protection of individual and collective rights in order to guarantee inclusion and social well-being. Besides, four social challenges are present to develop a national sustainable and inclusive AI. ENIA also constitutes an AI Consulting Board, which proposes initiatives to ensure safe and ethical AI use and development.



[Download here.](#)

14. Saudi Arabia, Saudi Data and Artificial Intelligence Authority:

SDAIA's transformation strategy was established in 2019. The strategy gives SDAIA a core mandate to drive and own the national data and AI agenda to help achieve Vision 2030's goals and Saudi's highest potential. To fulfill this mandate, SDAIA – and its sub-entities National Data Management Office NDMO (Legislative Arm), National Information Centre NIC (Operational Arm), and National Center for AI, NCAI (Creative Arm)– will deliver on the promise to create a data-driven and AI-supported government and economy.

[Download here](#)

15. Saudi Arabia, National Strategy for Data and AI NSDAI

Saudi Data & AI Authority (SDAIA) was established to drive the national agenda for Data & AI to elevate Saudi Arabia as a global leader in the elite league of data driven economies. To capitalize on Data & AI for the Kingdom economically and socially through national

combined efforts by all stakeholders, SDAIA has developed the National Strategy for Data & AI.

The strategy approval by King Salman bin Abdulaziz Al Saud was issued on July 17, 2020. SDAIA is currently working with different entities on launching the strategy and activating its initiatives. As the vision is "Where the best of Data & AI is made reality on", the objectives of NSDAI is 1. position KSA as a global data hub. 2. Transform KSA's workforce with a steady local supply of Data and AI-empowered talents. 3. Enact the most welcoming legislation for Data and AI businesses and talents. 4. Attract efficient, stable funding for qualified Data and AI investment opportunities. 5. Empower top Data & AI institutions to spearhead innovation and impact creation. 6. Stimulate Data and AI adoption with the most collaborative, and forward-thinking ecosystem.

[Download here](#)

16. United Arab Emirates: Strategy for Artificial Intelligence

In October 2017, the UAE Government launched the 'UAE Strategy for Artificial Intelligence (AI)'. This marks the post-mobile government phase which will rely on various future services, sectors and infrastructure projects. The strategy aims to achieve the objectives of UAE Centennial 2071, boost government performance at all levels, use an integrated smart digital system that can overcome challenges and provide quick efficient solutions, make the UAE the first in the field of AI investments in various sectors, and create new vital market with high economic value.

The AI strategy has 5 themes: (1) the formation of the UAE AI Council, (2) workshops, programmes, initiatives and field visits to government bodies, (3) develop capabilities and skills of all staff operating in the field of technology and organise training courses for government officials, (4) provide all services via AI and the full integration of AI into medical and security services, and (5) launch leadership strategy and issue a government law on the safe use of AI. The strategy will cover the following sectors: transport, health, space, renewable energy, water, technology, education, environment, and traffic.

[Available here](#)

17. United States: National AI R&D Strategic Plan: 2019 Update

National AI R&D Strategic Plan: 2019 Update identifies the critical areas of AI R&D that require Federal investments. Released by the White House Office of Science and Technology Policy's National Science and Technology Council, the Plan defines several key areas of priority focus for the Federal agencies that invest in AI. These areas of strategic AI R&D focus include: continued long-term investments in AI; effective methods for human-AI collaboration; understanding and addressing the ethical, legal, and societal implications for AI; ensuring the safety and security of AI; developing shared public datasets and environments for AI training and testing; measuring and evaluating AI technologies through standards and benchmark; better understanding the National AI R&D workforce needs; and expanding public-private partnerships to accelerate AI advances.

In September 2019, agencies for the first time reported their nondefense R&D investments in AI according to this Plan, through the [NITRD Supplement to the President's FY2020 Budget](#). This new AI R&D reporting process provides an important mechanism and baseline for consistently tracking America's prioritization of AI R&D going forward. This report also provides insight into the diverse and extensive range of nondefense Federal AI R&D programs and initiatives.

[Download here](#)

18. United States: Maintaining American Leadership in Artificial Intelligence - Executive Order 13859 of February 11, 2019

This Executive Order 13859 is signed by President Donald Trump on February 11, 2019 to announce the American AI Initiative — the United States' national strategy on artificial intelligence. This strategy is a concerted effort to promote and protect national AI technology and innovation. The Initiative implements a whole-of-government strategy in collaboration and engagement with the private sector, academia, the public, and like-minded international partners. This initiative takes a multipronged approach to accelerating AI development in US, and includes five key areas of emphasis:

(1) Investing in AI Research and Development (R&D)

The initiative focuses on maintaining strong, long-term emphasis on high-reward, fundamental R&D in AI by directing Federal agencies to prioritize AI investments in their R&D missions.

(2) Unleashing AI Resources

The initiative directs agencies to make Federal data, models, and computing resources more available to America's AI R&D experts, researchers, and industries to foster public trust and increase the value of these resources to AI R&D experts, while maintaining the safety, security, civil liberties, privacy, and confidentiality protections.

(3) Setting AI Governance Standards

Federal agencies will foster public trust in AI systems by establishing guidance for AI development and use across different types of technology and industrial sectors. This initiative also calls for the National Institute of Standards and Technology (NIST) to lead the development of appropriate technical standards for reliable, robust, trustworthy, secure, portable, and interoperable AI systems.

(4) Building the AI Workforce

This Initiative calls for agencies to prioritize fellowship and training programs to help American workers gain AI-relevant skills through apprenticeships, skills programs, fellowships, and education in computer science and other growing Science, Technology, Engineering, and Math (STEM) fields.

(5) International Engagement and Protecting our AI Advantage

[Download here](#)

19. United States: Guidance for Regulation of Artificial Intelligence Applications

Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence,"¹ requires the Director of the Office of Management and Budget (OMB), in coordination

with the Director of the Office of Science and Technology Policy, the Director of the Domestic Policy Council, and the Director of the National Economic Council, to issue a memorandum that provides guidance to all Federal agencies to inform the development of regulatory and non-regulatory approaches regarding technologies and industrial sectors that are empowered or enabled by AI and consider ways to reduce barriers to the development and adoption of AI technologies. Consistent with Executive Order 13859, the memorandum on *Guidance for Regulation of Artificial Intelligence Applications* was released in January 2020.

OMB guidance on these matters seeks to support the US approach to free markets, federalism, and good regulatory practices, which has led to a robust innovation ecosystem. When considering regulations or policies related to AI applications, agencies should continue to promote advancements in technology and innovation, while protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property.

[Download here](#)

International Organizations

1. Council of Europe: Unboxing AI: 10 steps to protect human rights

The Recommendation of the Commissioner for Human Rights to member states, *Unboxing AI: 10 steps to protect human rights*, was issued in May 2019. The Recommendation builds on work done in this area by the Council of Europe, in particular the European Ethical Charter on the use of artificial intelligence in judicial systems, the Guidelines on Artificial Intelligence and Data Protection, and the Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes.

The document identifies actionable points for 10 key areas, providing general guidance on operationalising the recommendations: (1) Human rights impact assessment; (2) Public consultations; (3) Obligation of member states to facilitate the implementation of human rights standards in the private sector; (4) Information and transparency; (5) Independent oversight; (6) Non-discrimination and equality; (7) Data protection and privacy; (8) Freedom of expression, freedom of assembly and association, and the right to work; (9) Remedies; (10) Promotion of “AI literacy”.

[Download here](#)

2. Council of Europe: Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems

The *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems* was adopted on 8 April 2020. The recommendation notes that, while algorithmic processes can foster innovation and economic development, they can also pose challenges to human rights, in areas such as privacy and data protection, freedom of thought, freedom of expression, the right to equal treatment, and economic and social rights.

The document sets out 9 recommendations for the governments of member states, which include establishing effective and predictable legislative and regulatory frameworks that prevent, detect, prohibit, and remedy human rights violations in the context of algorithmic systems, and ensuring that algorithmic systems incorporate safety, privacy, data protection, and security safeguards by design.

[Download here](#)

3. European Union: Communication Artificial Intelligence for Europe

In April 2018, the European Commission published its agenda for promoting artificial intelligence in Europe, *Artificial Intelligence for Europe*. This document sets out a European initiative on AI, with three strategic goals:

(1) Boost the EU's technological and industrial capacity and AI uptake across the economy, both by the private and public sectors. This includes investments in research and innovation and better access to data.

(2) Prepare for socio-economic changes brought about by AI by encouraging the modernization of education and training systems, nurturing talent, anticipating changes in the labor market, supporting labor market transitions and adaptation of social protection systems.

(3) Ensure an appropriate ethical and legal framework, based on the Union's values and in line with the Charter of Fundamental Rights of the EU. This includes forthcoming guidance on existing product liability rules, a detailed analysis of emerging challenges, and cooperation with stakeholders, through a European AI Alliance, for the development of AI ethics guidelines.

This document also points out that the European Commission will engage more member states and stakeholders to create and operate a broad multi-stakeholder platform, the European AI Alliance, to work on all aspects of AI.

[Download here](#)

4. European Union: Coordinated Plan on Artificial Intelligence

Cooperating with High-Level Expert Group on AI – a network of leading European AI experts-and the European AI Alliance, the European Commission published the Coordinated Plan on Artificial Intelligence in December 2018. Drawn up in collaboration with the Member States, the ambition is for Europe to become the world-leading region for developing and deploying cutting-edge, ethical and secure AI, promoting a human-centric approach in the global context. EU aims to increase investment and reach a total (public and private sectors combined) of at least EUR 20 billion in the period 2018-2020, and to increase investments progressively to EUR 20 billion per year in the course of the next decade. This plan details actions in 2019-2020 and prepares the ground for activities in the following years. It will be reviewed and updated annually.

The Communication highlights the main objectives and initiatives of the plan:

(1) Common objectives and complementary efforts. All Member States are encouraged to develop their national AI strategy by mid-2019, building on the work done at European level. (2) Towards a European AI public-private partnership and more financing for start-ups and innovative small and medium-sized enterprises. (3) Strengthening excellence in

trustworthy AI technologies and broad diffusion. (4) Adapting our learning and training programs and systems to better prepare society for AI. (5) Building up the European data space essential for AI in Europe, including for public sector. (6) Developing ethics guidelines with a global perspective and ensuring an innovation- friendly legal framework. (7) Security-related aspects of AI applications and infrastructure, and international security agenda.

The plan further points out eight fields for action: (1) Strategic actions and coordination; (2) Maximizing investments through partnerships; (3) From the lab to the market; (4) Skills and life-long learning; (5) Data: a cornerstone for AI – creating a Common European Data Space; (6) Ethics by design and regulatory framework; (7) AI for the Public Sector, and (8) International cooperation.

[Download here](#)

5. European Union: White Paper on Artificial Intelligence

Recognizing that the EU must act as one and define its own way, based on European values, to address the opportunities and challenges of AI, the European Commission published the *White Paper on Artificial Intelligence—A European approach to excellence and trust* in February 2020. The Commission supports a regulatory and investment-oriented approach with the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology.

The White Paper proposes measures that will streamline research, foster collaboration between Member States and increase investment into AI development and deployment, as well as policy options for a future EU regulatory framework that would determine the types of legal requirements that would apply to relevant actors, with a particular focus on high-risk applications.

[Download here](#)

6. Inter-American Development Bank (IDB) AI Strategy: fAIr LAC

The Inter-American Development Bank (IDB) aims to achieve a shared understanding of AI in LAC and how it can evolve in the near future, along with the opportunities it offers, its sectoral applications, and the risks it presents and the potential measures to mitigate that risk.

To that end, the IDB, in cooperation with several strategic partners, is leading an initiative called fAIr LAC with the aim of promoting the responsible development and application of AI to improve the delivery of services – thus reducing existing disparities and eventually reduce growing inequalities. The fAIr LAC initiative leverages the public and private sectors, as well as civil society, in order to make an impact on both public policy and the entrepreneurial ecosystem. The initiative was designed by the IDB in collaboration with C Minds, which received input from a diverse group of experts in the region. This document introduces the strategic definition and lines of action that fAIr LAC proposes in order to tackle the challenges and seize the opportunities that AI has to offer to societies of Latin America and the Caribbean.

The fAIr LAC initiative adopts the OECD's definition of AI, and its lines of action will center on any system that can create information for autonomous decision-making (ADM/S) in matters of social services delivery. This approach is independent from the type of learning and the type of algorithm, since AI model results constitute a data creation process that can interact in different ways with individuals responsible for making decisions to affect the environment, regardless of whether its result is introduced as a recommendation or if the system is capable of initiating an interim and final action. Nonetheless, the fact that fAIr LAC uses a broad definition of what constitutes AI, the fAIr LAC initiative will mainly center on risk assessment and the creation of sub-symbolic AI mitigation mechanisms, specifically for machine learning algorithms (Annex I), since they have become so popular in recent years due to their potential to improve the delivery of social services.

Given the challenges that have been identified and the scope determined for fAIr LAC, three strategic dimensions have been suggested that articulate strategies to achieve the objective to promote the responsible use of AI in the provision of social services. These three dimensions are (I) Development of a diverse network, (II) Training for responsible AI adoption, and (III) The promotion of quality and risk mitigation. Within each strategic dimension, lines of action are set forth with activities and specific interventions targeted at one or several of the identified users or beneficiaries, namely government, civil society organizations, enterprises, and citizenship.

[Download here](#)

7. G7: Charlevoix Common Vision for the Future of AI

The Vision, released in January 2019, follows the *2018 G7 Montreal Ministerial Statement on Artificial Intelligence* and the *2017 G7 ICT and Industry Ministers' Torino Declaration*, both of which advocate for a multi-stakeholder, human-centric approach.

The Vision highlights the commitments of G7 Leaders regarding human-centricity, personal data and privacy protection, inclusivity and empowerment of women and marginalized communities, accountability, assurance, liability, security, safety, multi-stakeholder dialogue, and transparency.

[Download here](#)

8. G20 Ministerial Statement on Trade and Digital Economy

The *G20 Ministerial Statement on Trade and Digital Economy* was released following the meeting of the G20 Trade Ministers and Digital Economy Ministers in Japan in June 2019. This document, to which the G20 AI Principles are annexed, is a result of a discussion on how digital policies could be designed to maximize benefits and minimize the challenges from the development of the digital economy, with special attention to developing countries and underrepresented populations.

The document advocates for human-centered AI, guided by the G20 AI Principles drawn from the OECD Recommendation on AI, and offers guidance for consideration by policy makers with the purpose of maximizing and sharing the benefits from AI, while minimizing the risks and concerns, with special attention to international cooperation and inclusion of developing countries and underrepresented populations.

[Download here](#)

9. OECD AI Policy Observatory: OECD.AI

Launched in February 2020, the Observatory aims to help policy makers and other stakeholders to implement the OECD AI Principles, through its activities related to: i) focused guidance on the AI Principles and their implementation; ii) analysis on AI in diverse policy areas from agriculture and development to skills and SMEs; iii) trends and data on AI developments to anchor policies in evidence; and iv) data on national AI policies and stakeholder initiatives.

[Download here](#)

10. UNOHR Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment.

Algorithms and Artificial Intelligence (AI) applications are now a critical part of the information environment – they are found in every corner of the internet, on digital devices and in technical systems, in search engines, social media platforms, messaging applications, and public information mechanisms. In this August 2018 report, former Special Rapporteur David Kaye examines the impact of AI on the information environment and proposes a human rights framework for the design and use of technologies comprising AI by states and private actors.

[Download here](#)

11. UNESCO: Beijing Consensus on AI and Education

The *Beijing Consensus on AI and Education* was published following the International Conference on Artificial Intelligence and Education held in Beijing from 16 to 18 May 2019. This first-ever document, developed by UNESCO in May 2019, to offer guidance on how to harness AI technologies for achieving the Education 2030 Agenda was adopted by over 50 government ministers, international representatives from over 105 Member States and almost 100 representatives from UN agencies, academic institutions, civil society and the private sector.

The consensus reaffirms a humanistic approach to the use of AI with a view towards protecting human rights and preparing all people with the appropriate values and skills needed for effective human-machine collaboration in life, learning and work, and for sustainable development. It advocates for human-controlled and human-centered AI development, where the deployment of AI should be in the service of people and to enhance human capacities; that AI should be designed in an ethical, non-discriminatory, equitable, transparent and auditable manner; and that the impact of AI on people and society should be monitored and evaluated throughout the value chains.

Based on these general principles, the consensus lays out concrete recommendations in the fields of planning in education policies; education management and delivery, empowering teachers; learning and learning assessment; and development of values and skills for life and work.

[Download here](#)

12. UNESCO: Steering AI and Advanced ICTs for Knowledge Societies: A ROAM Perspective

There are no simple answers about what the future holds for humanity, this 2019 report is a contribution to the wider debate about the ethics and governance of AI. It is an attempt to ‘steer’ clear of both technological utopianism, and dystopian thinking. Instead of technological determinism and its implication of inevitability, UNESCO gives attention to the role of human agency and human-centered values in the development of AI and other advanced information and communication technologies (ICTs).

This study frames its assessment of AI through UNESCO’s Internet Universality ROAM framework agreed by our Member States in 2015.

This study offers a set of options for action that can serve as inspiration for the development of new ethical policy frameworks and other actions, whether by States in their different fields of work, diverse actors in the private sector, members of academia and the technical community, and civil society.

[Download here](#)

13. UNESCO: I’d Blush If I Could

This publication by UNESCO on May 2019, advocates for gender equality in the field of AI, through closing the gender gap in digital skills and ending the perpetuation of gender stereotypes through female AI voice assistants.

The publication provides evidence for the gender gap; for example, in numerous countries, women are 25 per cent less likely than men to know how to leverage ICT for basic purposes. It also describes several reasons why this is detrimental to women, including compromising their safety; preventing them from reaping economic benefits; and limiting their ability to engage socially and politically. In response, it recommends increasing girls’ and women’s digital skills through early, varied and sustained exposure to digital technologies through formal and information education.

Simultaneously, the publication identifies that most AI voice assistants are female, causing reinforcement of gender bias; normalization of tolerance of sexual harassment and verbal abuse; and portraying women as the voices of servility and dumb mistakes. Potential solutions suggested including women in the process of developing AI, applying gender-sensitive approaches to digital skills development and creating new tools, processes and ensuring oversight.

[Download here](#)

14. UNESCO: AI Animation Series: 7 minutes to Understand AI and the key Human Rights and Ethical Principles Applying to its Development

This animation series is based on UNESCO’s publication Steering AI and Advanced ICTs for Knowledge Societies- a Rights, Openness, Access, and Multi-stakeholder (ROAM-X) Perspective and aims at promoting a better understanding of guiding human-centered and ethical principles underpinning any development of AI. The animations use everyday

language to explain and raise awareness on the key issues relating to AI development that might affect the achievement of SDGs. The animation series consists of seven episodes, each of which is about one minute long and illustrates a key message from the UNESCO's ROAM-X principles.

[Watch the animations here](#)

Civil society

1. Access Now: Human Rights in the Age of AI

Access Now's report proposes the use of international human rights law as a lens to examine AI and to provide solutions to some of the challenges that it poses, noting its advantages of having a system of institutions that provide well-developed frameworks for application of human rights to changing circumstances, and its normative power in the form of reputational and political costs. Interestingly, Access Now distinguishes human rights from ethics, acknowledging the role of ethical concepts while caveating that human rights are more universal and well-defined, and allow for accountability and redress.

The report describes some human rights which are impacted by AI, noting that vulnerable populations are often disproportionately impacted. The human rights examined are: rights to life, liberty, security, equality before the courts, a fair trial; rights to privacy and data protection; right to freedom of movement; rights to freedom of expression, thought, religion, assembly, and association; right to equality and non-discrimination; right to political participation and self-determination; prohibition on propaganda; rights to work, an adequate standard of living; right to health; right to education; right to take part in cultural life and enjoy benefits of scientific progress; right to marry, children's rights, and family rights; and right to life.

In conclusion, the report recommends that these human rights risks are examined; and that the principles of transparency, explainability, and accountability should concretely guide government and private sector use of AI.

[Download here](#)

2. Access Now: Privacy and Freedom of Expression in the Age of Artificial Intelligence (Privacy International & Article 19)

This paper focuses AI's impact on the right to privacy and the right to freedom of expression and information. It examines the ways in which AI impacts these two rights, reviews the landscape of AI governance, and provides some suggestions for rights-based solutions for civil society organizations and other stakeholders. Ultimately, the paper reiterates that compliance with human rights and regulatory standards should be a minimum requirement in the development and use of AI and that accountability and transparency is important for ensuring this compliance. It further suggests that civil society actors need to collect case studies of 'human rights critical' AI across the globe and actively engage in discussion with other stakeholders.


[Download here](#)

3. OECD: Network of Experts on AI: AI Repository of National AI Strategies and Policies

OECD.AI provides a live repository of over 600 national AI policy initiatives from over 60 countries, territories, and the EU. Building on the data contained in the repository, the OECD is developing practical guidance for policy makers on a wide array of topics: investing in AI R&D; data, infrastructure, software & knowledge; regulation, testbeds and documentation; skills and labour markets; and international co-operation, with the help of the OECD.AI Network of Experts.

[Download here](#)

4. W20: Artificial Intelligence: Open Questions About Gender Inclusion (W20)

The image shows the cover of a policy brief. At the top, it says 'W20' in a large, stylized font, with 'ARGENTINA 2018' and 'W20-18-02' below it. The title 'Artificial Intelligence: open questions about gender inclusion' is written in a green, sans-serif font. Below the title, there is a green abstract graphic with white circles and lines. At the bottom right, there is a small circular logo with the text 'WORLD LEADERS MEET' and 'W20'.

This policy brief published in October 2018 provides concrete recommendations to mitigate the challenges of AI related to gender (design, deployment, and collateral effects of digitalization strategies), with general principles embedded within them. Firstly, it suggests that countries need to take proactive steps to ensure that the process of designing AI technologies and policies are inclusive by including women in the AI workforce. Secondly, it proposes that women should be protected from discriminatory algorithms and that AI systems should be open and transparent to allow for monitoring. Lastly, it recommends that research should be conducted to assessment the effects of AI on women's lives.

[Download here](#)

5. Wonks and Techies: Input of the European Commission White Paper “On Artificial Intelligence- A European approach to excellence and trust”

Input on the European Commission White Paper “On Artificial Intelligence – A European approach to excellence and trust”

BY: “WONKS AND TECHIES,”
a multi-disciplinary group at Stanford University, cooperating on international technology and policy issues, led by Ms. Marietje Schaake

The multidisciplinary group at Stanford University, cooperating on international technology and policy issues, named Wonks and Techies published on June 2020 Input on the European Commission White Paper: “On Artificial Intelligence –A European approach to excellence and trust”. The key conclusion of the white paper includes: “We want to see the EU promote an equitable distribution of AI research, development and deployment. We encourage initiatives to increase public awareness, training, and literacy in response to advancements in AI, and suggest the creation of new occupations in the data-driven future. These recommendations can be coordinated and operationalized throughout the EU, made up of distinguished interdisciplinary experts, to tackle the implementation of dynamic policies as they relate to the development and trade of AI hardware and software, cooperation, and the capacity for change.”

[Download here](#)

Chapter 5: Observations on the ways forward

Artificial intelligence has captured great attention in recent years from governments, academia, and the public, amidst a growing recognition that AI will have broad impacts across societies, economies and the environment. But AI is not a new technology. The Turing Test for constructing a ‘thinking machine’ dates to a paper in 1950, and the term ‘Artificial Intelligence’ was first coined at the Dartmouth Summer Research Project in 1956. AI has undergone more than 60 years of development but has advanced more rapidly in recent years in step with technical advances in computational infrastructure, networked data and algorithmic innovation.

AI can be simply described as the ‘brain’ of a machine or a tool. It refers to computers imitating the functionalities of the human brain, such as perceiving, speaking, thinking, reasoning, and learning. AI technology has demonstrated the ability to surpass human ability in certain tasks. For instance, DeepMind Technologies developed an AI named AlphaGo that taught itself to master complex games like Go and to defeat the world Go champion, Lee Sedol. AI also has been integrated into most daily lives for a long time. Smartphone apps, web searches, spam filters, facial recognition, call center voice assistance and many other services already in wide use are examples of AI applications. For higher-level applications, AI is being applied in healthcare diagnoses, financial risk management, agricultural monitoring, autonomous vehicles, and many other purposes to be imagined and developed. AI is a complex and expanding field that requires multi-disciplinary input and wider public understanding.

The accelerating development of AI provides numerous positive benefits to advance the achievement of Sustainable Development Goals. AI can improve the effectiveness of public services and enhance government accountability and transparency. AI has created significant possibilities for economic efficiencies and productivity growth. In manufacturing, for instance, AI drives the application of robotics and autonomous networked technologies that reduce production time and costs.

Frontier technologies such as the Internet of Things (IoT), blockchain and quantum computing do not exist as independent technologies from AI. Every technology plays a role in supporting the production of others, giving prominence to the collaboration between these technologies. IoT allows people to monitor production and other ongoing activities remotely, which can be utilized in marine management, agriculture and more. Combined use of IoT and AI can utilize data for effective control of sustainable production to reduce CO₂ emission. AI can make education and healthcare more accessible, which helps to reduce inequality in the long term. The climate crisis underscores the need to harness the full power of game-changing technologies like AI to make a decisive break

towards a new, more sustainable and equitable ‘normal’. Many further benefits are yet to be discovered.

The development and use of AI technologies will continue to transform society, which, in turn, will call for ethical reflection to guide how humans interpret their moral agency in relation to technological objects. Understanding the transitions that may take place in societies will also require multi-disciplinary research and multi-stakeholder cooperation, including for the periodic updating of ethical standards that guide the development and use of AI. The UN system provides a platform to facilitate standard setting, exchange of knowledge, and cooperation among different stakeholder groups from the global North and South.

Multi-stakeholder engagement, such as the UN’s Technology Facilitation Mechanism and its annual Global STI Forum, informed dialogue, and strengthening capacities, particularly in poorer countries, will be essential in such an endeavor.

Technical standards, such as specifying requirements on AI systems’ transparency, explainability, robustness, and interoperability between and among products will affect how AI is researched and developed. If standards are adopted on the international level, particularly through public policy, they can spread good practices, improve quality of life internationally, and facilitate global trade. AI technical standards can guide the development of AI in a safe and constructive manner.

There appears to be an increasing interest to review both policy and regulatory frameworks on AI, as well as existing national AI strategies and technical standards. This Resource Guide has demonstrated that many actors have taken on the challenge of defining principles, standards, and responsible uses of AI. The requirements of transparency, accountability, auditability, respect for human rights, data protection, equity, and technical robustness come up repeatedly throughout.

Different actors have emphasized distinctive areas of focus in their principles. For example, US entities have delved deeper into operational specificity while China has stressed the importance of international cooperation and open sharing of resources. EU legislation on AI will be formed based on the AI Ethics Guidelines from the AI High-Level Expert Group, which particularly accentuates diversity and non-discrimination.

The resources provided in this Guide articulate international, regional, and national approaches to the ethics and principles of AI development in order to provide an overview of global discussions. These discussions present a roadmap for cooperation toward ethically informed and human rights-based governance of AI. UNESCO’s standard setting process on the ethics of AI, among other initiatives like the UN Secretary General’s High-Level Panel on Digital Cooperation and ITU’s AI for Good Summit, work to facilitate cooperation on AI within the UN system and its Member States.

Despite a significant body of literature, there is a dearth of information and reliable analysis at the sectoral and country level, and many questions remain unanswered. Following this Resource Guide, UNDESA and its partners intend to review and propose a research agenda covering the main elements and key issues for policy development beyond the Guide's scope.

The need for AI comparable indicators

There is a need for methodological frameworks shared among countries to measure AI impacts among individuals and companies. Many Member States are facing increasingly pressing challenges for the production of more detailed, robust and timely data with scarce resources, particularly in the context of the COVID19 pandemic.

AI and other digital technologies can also be instrumental for data production in the context of the pandemic, adding further complexity to the ICT data production ecosystem as new data sources emerge. Such alternative sources and approaches for data production have played a crucial role in times of crisis and are of utmost importance for informed decision-making by citizens, businesses and governments.

Measuring AI

There is ongoing investigation led by the OECD through the “Working Party on Measurement and Analysis of the Digital Economy – WP MADE” on how to measure AI using ICT usage surveys. One example is the paper “AI measurement in ICT usage surveys: a review” that takes stock of official statistics on AI use in firms collected through ICT usage surveys. Its aim is to highlight statistically sound data that can be used to guide policymakers and other stakeholders in the complex field of AI. The document has been discussed in the working group has not yet been made public yet.

As AI continues to evolve and grow, governance will need to keep pace of trends on where we are heading in order to develop a safe, equitable, and beneficial path for AI implementation in the future. A multidisciplinary, multistakeholder approach will assist in addressing ethical challenges and ultimately contribute to achieving the 17 Sustainable Development Goals.³⁰

³⁰ Citations for this chapter:

Benaich, N., & Hogarth, I. (2020, October 01). State of AI Report 2020. Retrieved from <https://www.stateof.ai/>

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., ... & Trench, M. (2017). Artificial intelligence: The next digital frontier. McKinsey Global Institute, 1-80. Retrieved from <https://www.mckinsey.com/~media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>

Bughin, J., Manyika, J., Catlin, T., (2019, May 21). Twenty-Five Years Of Digitization: Ten Insights Into How To Play It Right. McKinsey Global Institute. Retrieved from <https://www.mckinsey.com/business->

-
- functions/mckinsey-digital/our-insights/twenty-five-years-of-digitization-ten-insights-into-how-to-play-it-right.
- Burke, B. (2020, September 16). Coronavirus updates (Sept. 14-Sept. 28): Coronavirus effects on private markets. Retrieved from <https://pitchbook.com/news/articles/coronavirus-updates-latest-news-and-analysis-september-14-september-28>.
- Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018, April 17). Notes from the AI frontier: Applications and Value Of Deep Learning. McKinsey Global Institute. Retrieved from <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>.
- COVID Moonshot. (n.d.). Retrieved from <https://covid.postera.ai/covid>
- Disengagement Reports. (2020, June 06). Retrieved from <https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/disengagement-reports/>
- Elmi, N., Broekaert, K., & Larsen, A. M. E. (2018, January). Agile Governance: Reimagining Policy-Making in the Fourth Industrial Revolution. In White Paper. World Economic Forum, January. Retrieved from <https://www.weforum.org/whitepapers/agile-governance-reimagining-policy-making-in-the-fourth-industrial-revolution>
- Frontier technologies for sustainable development in Asia and the Pacific. (2018, May 3). UNESCAP. Retrieved from <https://www.unescap.org/publications/frontier-technologies-sustainable-development-asia-and-pacific>.
- Ghosh, I. (2020, May 22). Mapped: The State of Facial Recognition Around the World. Retrieved from https://www.visualcapitalist.com/facial-recognition-world-map/?utm_source=morning_brew
- Goasduff, L. (2020, September 28). 2 Megatrends Dominate the Gartner Hype Cycle for Artificial Intelligence, 2020. Retrieved from <https://www.gartner.com/smarterwithgartner/2-megatrends-dominate-the-gartner-hype-cycle-for-artificial-intelligence-2020/>
- Gofman, M., & Jin, Z. (2020, October 26). *Artificial Intelligence, Education, and Entrepreneurship*. Retrieved from http://gofman.info/AI/AI_GofmanJin.pdf
- Hill, K. (2020, January 18). The Secretive Company That Might End Privacy as We Know It. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>
- Papers With Code Trends. (n.d.). Retrieved from <https://paperswithcode.com/trends>
- Pyle, D., & San Jose, C. (2015, June 01). An Executive's Guide to Machine Learning, McKinsey Quarterly: June, 2015. Retrieved from <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/an-executives-guide-to-machine-learning>
- Sicat, M.. (2018, April 18). Overview: Emerging Technologies and Issues, Geneva CSTD. Retrieved from https://unctad.org/meetings/en/Presentation/dtl_eWeek2018p60_MarieSicat_en.pdf.
- Sirimanne, S., Bell, B., Fajarnés, P., Sanz, A., Lim, M., Ok, T., . . . Ting, B. (2018). Technology and Innovation Report 2018: Harnessing Frontier Technologies for Sustainable Development. UNCTAD. Retrieved from <https://unctad.org/en/pages/publications/Technology-Innovation-Report.aspx>.
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., & Teller, A. (2016, September). One Hundred Year Study on Artificial Intelligence. Artificial Intelligence and Life in 2030. Retrieved from <https://ai100.stanford.edu/2016-report>
- Sumitomo Dainippon Pharma Co., Ltd and Exscientia Ltd. (2020, January 30). *Sumitomo Dainippon Pharma and Exscientia Joint Development New Drug Candidate Created Using Artificial Intelligence (AI) Begins Clinical Study* [Press release]. Retrieved from <https://www.ds-pharma.com/ir/news/pdf/ene20200130.pdf>
- The Global AI Talent Tracker. (2020, June 10). Retrieved from <https://macropolo.org/digital-projects/the-global-ai-talent-tracker/>
- World Economic Forum. (2018, January). The next economic growth engine: Scaling Fourth Industrial Revolution technologies in production. The World Economic Forum in collaboration with.

Bibliography

- Artificial intelligence. (2020, September 02). National Institute of Standards and Technology. Retrieved from <https://www.nist.gov/artificial-intelligence>
- Benaich, N., & Hogarth, I. (2020, October 01). State of AI Report 2020. Retrieved from <https://www.stateof.ai/>
- Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., ... & Trench, M. (2017). Artificial intelligence: The next digital frontier. McKinsey Global Institute, 1-80. Retrieved from <https://www.mckinsey.com/~media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>
- Bughin, J., Manyika, J., Catlin, T., (2019, May 21). Twenty-Five Years Of Digitization: Ten Insights Into How To Play It Right. McKinsey Global Institute. Retrieved from <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/twenty-five-years-of-digitization-ten-insights-into-how-to-play-it-right>.
- Burke, B. (2020, September 16). Coronavirus updates (Sept. 14-Sept. 28): Coronavirus effects on private markets. Retrieved from <https://pitchbook.com/news/articles/coronavirus-updates-latest-news-and-analysis-september-14-september-28>.
- Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018, April 17). Notes from the AI frontier: Applications and Value Of Deep Learning. McKinsey Global Institute. Retrieved from <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>.
- COVID Moonshot. (n.d.). Retrieved from <https://covid.postera.ai/covid>
- Disengagement Reports. (2020, June 06). Retrieved from <https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/disengagement-reports/>
- ECOSOC High-level Political Forum on Sustainable Development. (2019, May 29). *Multi-stakeholder forum on science, technology and innovation for the Sustainable Development Goals: summary by the co-chairs*. Retrieved from https://www.un.org/ga/search/view_doc.asp?symbol=E/HLPF/2019/6&Lang=en
- Elmi, N., Broekaert, K., & Larsen, A. M. E. (2018, January). Agile Governance: Reimagining Policy-Making in the Fourth Industrial Revolution. In White Paper. World Economic Forum, January. Retrieved from

- <https://www.weforum.org/whitepapers/agile-governance-reimagining-policy-making-in-the-fourth-industrial-revolution>
- Frontier technologies for sustainable development in Asia and the Pacific. (2018, May 3). UNESCAP. Retrieved from <https://www.unescap.org/publications/frontier-technologies-sustainable-development-asia-and-pacific>.
- Ghosh, I. (2020, May 22). Mapped: The State of Facial Recognition Around the World. Retrieved from https://www.visualcapitalist.com/facial-recognition-world-map/?utm_source=morning_brew
- Goasduff, L. (2020, September 28). 2 Megatrends Dominate the Gartner Hype Cycle for Artificial Intelligence, 2020. Retrieved from <https://www.gartner.com/smarterwithgartner/2-megatrends-dominate-the-gartner-hype-cycle-for-artificial-intelligence-2020/>
- Gofman, M., & Jin, Z. (2020, October 26). *Artificial Intelligence, Education, and Entrepreneurship*. Retrieved from http://gofman.info/AI/AI_GofmanJin.pdf
- Hill, K. (2020, January 18). The Secretive Company That Might End Privacy as We Know It. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>
- Hu, X., Neupane, B., Echaiz, L. F., Sibal, P., & Rivera Lam, M. (2019). *Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective*. UNESCO Publishing. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000372132>
- McDonald, H. (2020, August 04). Home Office to scrap 'racist algorithm' for UK visa applicants. *The Guardian*. Retrieved from <https://www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racist-algorithm-for-uk-visa-applicants>
- McDonald, H. (2020, August 04). Home Office to scrap 'racist algorithm' for UK visa applicants. *The Guardian*. Retrieved from <https://www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racist-algorithm-for-uk-visa-applicants>
- Microsoft deletes massive face recognition database. (2019, June 07). *BBC News*. Retrieved from <https://www.bbc.com/news/technology-48555149>
- Microsoft deletes massive face recognition database. (2019, June 07). *BBC News*. Retrieved from <https://www.bbc.com/news/technology-48555149>
- New Zealand. (2020, July). *Algorithm Charter for Aotearoa New Zealand*. Retrieved from https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

- New Zealand. (2020, July). *Algorithm Charter for Aotearoa New Zealand*. Retrieved from https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf
- Papers With Code Trends. (n.d.). Retrieved from <https://paperswithcode.com/trends>
- Pyle, D., & San Jose, C. (2015, June 01). An Executive's Guide to Machine Learning, McKinsey Quarterly: June, 2015. Retrieved from <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/an-executives-guide-to-machine-learning>
- Schlosser, M. (2019, October 28). Agency. *The Stanford Encyclopedia of Philosophy* (Winter 2019). Retrieved from <https://plato.stanford.edu/entries/agency/>
- Sicat, M.. (2018, April 18). Overview: Emerging Technologies and Issues, Geneva CSTD. Retrieved from https://unctad.org/meetings/en/Presentation/dtl_eWeek2018p60_MarieSica_t_en.pdf.
- Sirimanne, S., Bell, B., Fajarnés, P., Sanz, A., Lim, M., Ok, T., . . . Ting, B. (2018). Technology and Innovation Report 2018: Harnessing Frontier Technologies for Sustainable Development. UNCTAD. Retrieved from <https://unctad.org/en/pages/publications/Technology-Innovation-Report.aspx>.
- Special Address by Antonio Guterres, Secretary-General of the United Nations. (2020, January 23). *World Economic Forum*. Retrieved from <https://www.weforum.org/events/world-economic-forum-annual-meeting-2020/sessions/special-address-by-antonio-guterres-secretary-general-of-the-united-nations-1>
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., & Teller, A. (2016, September). One Hundred Year Study on Artificial Intelligence. Artificial Intelligence and Life in 2030. Retrieved from <https://ai100.stanford.edu/2016-report>
- Sumitomo Dainippon Pharma Co., Ltd and Exscientia Ltd. (2020, January 30). *Sumitomo Dainippon Pharma and Exscientia Joint Development New Drug Candidate Created Using Artificial Intelligence (AI) Begins Clinical Study* [Press release]. Retrieved from <https://www.ds-pharma.com/ir/news/pdf/ene20200130.pdf>
- The Global AI Talent Tracker. (2020, June 10). Retrieved from <https://macropolo.org/digital-projects/the-global-ai-talent-tracker/>
- UNDESA. (2019, May 14). Session 1: Emerging Technology Clusters and The Impact Of Rapid Technological Change On The SDGs. *Sustainable Development Knowledge Platform*. Retrieved from

- <https://sustainabledevelopment.un.org/index.php?page=view&type=20000&nr=5516&menu=2993> .
- UNDESA. (n.d.). Technology Facilitation Mechanism Workstream 10: Analytical work on emerging technologies and the SDGs. *Sustainable Development Knowledge Platform*. Retrieved from <https://sustainabledevelopment.un.org/index.php?page=view&type=12&nr=3335>
- UNECE. (2018, September 26). Standards for the Sustainable Development Goals, Geneva CIG. Retrieved from <https://www.unece.org/sdgs-isoweek2018.html>
- UNESCO. (2020). Records of the General Conference, 40th session, Paris, 12 November-27 November 2019, *volume 1: Resolutions*. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000372579.nameddest=37>
- UNSG's High-level Panel on Digital Cooperation. (2019). *The Age of Digital Interdependence*. Retrieved from <https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>
- Weise, K., & Singer, N. (2020, June 10). Amazon Pauses Police Use of Its Facial Recognition Software. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/06/10/technology/amazon-facial-recognition-backlash.html>
- Weise, K., & Singer, N. (2020, June 10). Amazon Pauses Police Use of Its Facial Recognition Software. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/06/10/technology/amazon-facial-recognition-backlash.html>
- World Commission on the Ethics of Scientific Knowledge and Technology. (2017). *Report of COMEST on robotics ethics*. UNESCO. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000253952>
- World Economic Forum. (2018, January). The next economic growth engine: Scaling Fourth Industrial Revolution technologies in production. The World Economic Forum in collaboration with McKinsey & Company. Retrieved from <https://www.weforum.org/whitepapers/the-next-economic-growth-engine-scaling-fourth-industrial-revolution-technologies-in-production>.
- WSIS. (2003, December 12). Declaration of Principles: Building the Information Society: a Global Challenge in the New Millennium. *World summit on the information society*. Retrieved from <https://www.itu.int/net/wsis/docs/geneva/official/dop.html>

Annex I: Roadmap for the UNESCO Recommendations on the Ethics of Artificial Intelligence

TIMEFRAME	ROADMAP ACTION
April 2020	First meeting of the 24-member Ad Hoc Expert Group convened to prepare the draft text of the Recommendation.
July-August 2020	Global public online consultation and eleven regional and sub-regional virtual consultations. More than 600 submissions and 50,000 suggestions received on the first draft text.
August-September 2020	Second meeting of the Ad Hoc Expert Group to revise the draft text of the Recommendation based on outcomes of consultations.
September 2020	Draft text of the Recommendation transmitted to Member States for written comments to be received by 31 December 2020.
January-March 2021	UNESCO Secretariat prepares a final report containing the draft text of the Recommendation based on comments and observations from Member States
April 2021	Transmission of the final report containing text of the Recommendation to Member States. First session of the special committee of intergovernmental experts to prepare a final draft of the Recommendation.
June 2021	Second session of the special committee of intergovernmental experts to prepare a final draft of the Recommendation.
Mid-August 2021	Transmission of the final draft of the Recommendation by the special committee of intergovernmental experts to Member States.
Autumn 2021	41st General Conference: Examination and possible adoption of the final draft Recommendation.

Annex II: Roadmap for the UNESCO Recommendation on the Ethics of Artificial Intelligence

ICAO

- UN AI for Good Annual Global Summit (expected date: 21-25 September 2020): presentation on AI in aviation, as well as participation in mobility sessions. AI for Good is a United Nations platform led by the International Telecommunication Union (ITU) and the XPrize Foundation, which considers AI as an accelerator for the realization of the UN Sustainable Development Goals (SDGs). The third edition of the AI for Good Annual Global Summit in 2020 led to the launch of the Global Initiative on AI and Data Commons, to which ICAO will be participating. The goal of this initiative is to develop an enabling infrastructure to help problem owners and AI innovators collaborate and accelerate progress towards the UN SDGs, knowing that solutions would be common resources.
- Development of in-house capabilities: hosting internships, developing ICAO deep learning AI models showcasing natural language processing techniques for aeronautical information management, as well as document summarization.
- Support to local AI company networks and start-ups and incubators such as Thales AI@Centech and Concordia District 3, by providing ideas and coaching. This approach allows ICAO to have a direct link to applicable AI solutions and to collaboratively approach the underlying challenges of AI and ethics.
- Collaborating with McGill University on the introduction of AI in aviation course inside the McGill data science and AI program: AI and ethics should be a module under this introductory course.
- Broadening the horizon of AI in aviation through workshops organized in collaboration with the Consortium de recherche et d'innovation en aérospatiale au Québec (CRIAQ): ICAO will require the consideration of ethical aspects of AI as part of the workshops' agendas.
- Exploring the creation of an AI in Aviation Focus Group under the auspices of ITU to address issues related to compliance and certification, knowing that components on ethics could also be part of the compliance aspects
- **ILO research program on Technologies and the Future of Work** addresses the impact of technology, including artificial intelligence (AI) on jobs, employment, decent work, productivity, inequality and sustainable development.
- **The Report of the Global Commission on the Future of Work “Work for a brighter future”** (January 2019) subscribes to a “human-in-command” approach to AI that ensures that the final decisions affecting work are taken

by human beings, not algorithms. It also calls for the establishment of “an international governance system for digital labour platforms”.

IOM

- IOM leads **an inter-agency group on Data Science, Artificial Intelligence and Ethics**, which established inter-agency peer review mechanisms for mathematical AI models and ethics.
- IOM co-leads, with OCHA and UNHCR, the **IASC RG1 Sub-Group on Data Responsibility**, tasked with developing “Joint System-Wide Operational Guidance on Data Responsibility in Humanitarian Action”
- IOM co-organized with the German Federal Foreign Office (FFO) an **interagency workshop on “Forecasting Human Mobility in Contexts of Crises”**, touching on diverse aspects of data science, including machine learning and artificial intelligence.
- IOM funded the **“The Signal Code: Ethical Obligations for Humanitarian Information Activities”**, published by the Harvard Humanitarian Initiative in 2018.

ITU

- **The AI for Good Global Summit** organized by the ITU in partnership with 37 UN sister agencies, XPRIZE and ACM, seeks to ensure trusted, safe and inclusive development of AI technologies and equitable access to their benefits.
- **The ITU/WHO AI for Health Focus Group** serves as a benchmarking framework for AI-enabled healthcare solutions so that they can be deployed responsibly and in the right context of use for all.
- **The ITU-UNESCO Broadband Commission for Sustainable Development’s Working Group on AI and Global Health** facilitates advocacy efforts such as to generate knowledge on successes, challenges, and lessons learned from AI solutions in health.
- **The ITU Telecommunication Standardization Sector (ITU-T) Focus Group on AI for autonomous and assisted driving (FG-AI4AD)** supports standardisation activities of AI evaluation in autonomous and assisted driving.
- **The ITU-T Focus Group on Machine Learning for Future Networks including 5G** analyses the impact of ML on communication networks including studies on functional network architectures, data structures, interfaces, protocols and algorithms based on use cases from industry.
- The **ITU-T Focus Group on Environmental Efficiency for Artificial Intelligence and other Emerging Technologies** (FG-AI4EE) identifies the standardization needs to develop a sustainable approach to AI and other emerging technologies including automation, augmented reality, virtual reality, extended reality, smart manufacturing, industry 5.0, cloud/edge computing, nanotechnology, 5G, among others.

- AI acts as an enabler to enhance emerging radio technologies, and therefore is increasingly playing a part in the work of the ITU Radio Communications Sector (ITU-R) in activities related to managing frequency spectrum and satellite orbits, as well as developing globally harmonized regulations and standards, are fundamental in order to ensure accessible and affordable telecommunications to all. These include IMT2020 (5G), IoTs, Cognitive Radio Systems, Satellite Ecosystem, Spectrum Monitoring, Intelligent Transport System, and Broadcasting.

OHCHR

- **The OHCHR-UN Global Pulse Conference on a Human Rights-based approach to AI**
- The **High Commissioner for Human Rights’ Report on the Right to Privacy in the Digital Age** ([A/HRC/39/29](#)) addressed the rise of data-driven technologies and made recommendations for rights-protective measures.
- Special Rapporteur David Kaye’s Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment for the **promotion and protection of the right to freedom of opinion and expression** ([A/73/348](#))
- **An expert seminar on the impact of AI on the enjoyment of the right to privacy** will be organized in 2020, with a thematic report on this topic to the Human Rights Council in September.
- OHCHR works closely with the Advisory Committee of the Human Rights Council on addressing human rights-issues related to digital technology, including AI.
- OHCHR also provides input into the work of several treaty bodies concerning AI (e.g. the **draft General Recommendation on racial profiling** of the **Committee on the Elimination of Racial Discrimination** and the **draft General Comment on the right of peaceful assembly** of the **Human Rights Committee**).
- **B-Tech project on the application of the UN Guiding Principles on Business and Human Rights** to the development and use of digital technologies including AI.

UN DESA

- **The 2018 World Economic and Social Survey (WESS) on [Frontier Technologies for Sustainable Development](#)** analyzed (1) efficiency gains and equity and ethical concerns in relation to AI-based decision-making systems both in the public and private sector, and (2) production of targeted advertisements, manipulation of human emotion and spread of misinformation, including hatred.
- **A paper entitled [“Artificial Intelligence: Opportunities and Challenges for the Public Sector”](#)** addresses *“Ethical considerations for policy makers in the era of AI-centric approach”*

- **Global Working Group (GWG) on Big Data** has a task team on Privacy Preserving Techniques. This team produced a [UN Handbook on Privacy-Preserving Computation Techniques](#).
- [2018 United Nations E-Government Survey](#) Chapter 8 entitled “Fast-evolving technologies in e-government: Government Platforms, Artificial Intelligence and People” discusses transformative technologies, such as data analytics, artificial intelligence including cognitive analytics, robotics, bots, high-performance and quantum computing.
- **UN Technology Facilitation Mechanism (TFM)**
 - **Multi-Stakeholder Forum on Science, Technology and Innovation for the SDGs (“STI Forum”)** is the premier UN space for discussions on STI for the SDGs, including such cross-SDG issues such as emerging technologies and their sustainable development impact.
 - **Interagency Task Team on STI for the SDGs (IATT)**, through its work stream 10 (“Analytical work on emerging technologies and the SDGs”), have worked towards assessing the impacts of rapid technological change on the SDGs, including through UN expert group meetings to discuss the economic, societal and environmental impacts and ethical dimensions of artificial intelligence. Core principles and recommendations on responsible AI was suggested by experts in the contexts of the work under TFM and in particular the IATT’s subgroup on new and emerging technologies.
- **The Commission for Social Development** address [“Innovation and interconnectivity for social development”](#) as an emerging issue, while **“Socially just transition towards sustainable development: the role of digital technologies on social development and well-being of all”** will be a priority theme for its 59th session in 2021.
- Annual observance of the **International Day of Persons with Disabilities** on 3 December 2014 under the theme *“Sustainable Development: The promise of technology”*.
- A roundtable on **“Technology, digitalization and information and communications technology for the empowerment and inclusion of persons with disabilities”** was organized at the 12th session of the Conference of States Parties to the Convention on the Rights of Persons with Disabilities in 2019
- **DESA-ITU side event on “Why it Matters: AI for Older Persons”** (18 April 2019) at the 10th working session of the General Assembly’s open-ended working group for the purpose of strengthening the protection of the human rights of older persons.
- DESA will examine the potential for further research, including in collaboration with young researchers for 2021 and beyond to create a **youth research collaborative to investigate further the potential impacts of AI from a youth perspective.**
 - DESA will produce a **policy paper focusing on the potential socioeconomic impacts of digital technologies on with a particular**

focus on youth, given that they will experience much of the changes driven by AI

- The **2021 World Youth Report** has the theme **“Safe and Inclusive Digital Spaces for Youth”**, which will explore issues around online data management, disinformation, health and wellbeing, cybersecurity, and human rights etc. in the context of increasing youth engagement in digital spaces mediated by AI.

UN Women

- UN Women has launched the [Gender Innovation Principles](#) to set standards for a gender-responsive approach to innovation and technology, including AI. These principles could help guide private sector, academic and non-profit institutions to make sure women’s and men’s concerns and experience are equally integrated in the design of AI and machine learning systems and that due consideration is given to gender norms.
- UN Women published [“Leveraging Digital Finance for Gender Equality and Women’s Empowerment”](#) which identifies recommendations to mitigate the risk associated with AI and digital finance. This includes advocating for societal values to be reflected in algorithms and AI technologies and making sure that companies and people using algorithms can be held accountable and that their decisions are transparent. It also includes recommending companies to audit their algorithms to prevent the development of biased services.

UNCTAD

- **Technology and Innovation Report (TIR)**
 - **TIR 2018 “Harnessing Frontier Technologies for Sustainable Development”** explored how harnessing frontier technologies could be transformative in achieving the Sustainable Development Goals (SDGs).
 - **Forthcoming TIR 2020** will outline the state-of-the-art debate and critically examine the possibility of frontier technologies (including AI) widening existing inequalities and creating new ones.
- UN Secretary-General’s Report for the UN Commission on Science and Technology for Development (CSTD) on **“Harnessing rapid technological change for inclusive and sustainable development” (E/CN.16/2020/2)** and on **“Impact of rapid technological change on sustainable development” (E/CN.16/2019/2)** discussed the need for a consistent public policy response to the normative challenges posed by frontier technologies, notably Artificial Intelligence.
- Session entitled **“Structural transformation, Industry 4.0 and inequality: Science, technology and innovation policy challenges”**, at the Eleventh session of the Investment, Enterprise and Development Commission
- **Digital Economy Report 2019 on “Value Creation and Capture: Implications for Developing Countries”**, focused on the central role of data in economic processes. It analysed the evolution of the data-driven digital economy and

highlighted power imbalances and inequalities in access and use of data, as well as in the evolution of emerging technologies, including AI.

UNEP

- **Science Policy Business Forum of UNEP** has been holding discussions and consultations related to the implications around data and AI.
- **Partnership with Google Earth Engine and the EU JRC** to deploy machine-learning algorithms to detect global surface freshwater from open source satellite images as a baseline data set for indicator SDG 6.6.1.
- **Partnership with Global AI** on using document scraping techniques to assess the compliance of Corporate Sustainability Reports to certain standards.

UNESCO

- Culture, Platforms and Machines: the Impact of AI on the Diversity of Cultural Expressions- a report to the Intergovernmental Committee for the Protection and Promotion of the Diversity of Cultural Expressions
- I'd Blush If I Could- a report that advocates for gender equality in the field of AI
- AI Animation Series: 7 minutes to Understand AI and the key Human Rights and Ethical Principles Applying to its Development
- Beijing Consensus on AI and Education
- As a follow-up of World Summit on the Information Society's (WSIS), UNESCO **is** responsible taken responsibility for the implementation of the Action Lines on Access (C3), E-Learning (C7), Cultural diversity (C8), Media (C9), and Ethical dimension of the information society (C10).
- Member States of UNESCO adopted the **framework of "Internet Universality"** and the associated **"R.O.A.M. principles" (Human Rights, Openness, Accessibility and Multi-stakeholder participation) in 2015; the Sustainable Development Goals**. The 303 [Internet Universality ROAM-X Indicators](#) to assess how well national stakeholders, including governments, companies, and civil society perform in adhering to the ROAM principles were developed over a three-year process of global and inclusive consultations with stakeholders and was endorsed for voluntary national assessment in November 2018 by the 31st Council of the International Programme for the Development of Communication (IPDC). A publication entitled **"Steering AI and Advanced ICTs for Knowledge Societies: a ROAM perspective"** was launched at the Internet Governance Forum in 2019.
- UNESCO's World Commission on Ethics of Scientific Knowledge and Technology (COMEST) has prepared a **Preliminary Study on Ethics of Artificial Intelligence**, which triggered the decision of UNESCO Member States to elaborate a **Recommendation on the Ethics of Artificial Intelligence**.
- UNESCO has also organized a series of events addressing the ethical, legal and social implications of AI. Some of major events include:

- **Roundtables on "Artificial Intelligence: Reflection on its complexity and impact on our society"** (Paris, September 2018 & December 2019);
- **Workshop on "Artificial Intelligence for Human Rights and SDGs: Fostering Multi-Stakeholder, Inclusive and Open Approaches"** (Paris, November 2018);
- **Forum on artificial intelligence in Africa** (Ben Guérir, December 2018);
- **Debate on Ethics of New Technologies and Artificial Intelligence "Tech Futures: Hope or Fear?"** (Paris, January 2019);
- **UNESCO Conference "Principles for AI: Towards a Humanistic Approach?"** (March 2019);
- **International Conference on Artificial Intelligence and Education** (Beijing, May 2019), with **Beijing Consensus on Artificial Intelligence and Education** as the outcome document;
- **Youth Voices and the Future of Artificial Intelligence: Towards a Human-Centered Approach** (Paris, November 2019);
- **Regional Forum on AI in Latin America and the Caribbean** (Sao Paulo, December 2019)
- **Launch of the AI animation series "7 minutes to understand AI and key ROAM-X principles applying to its development"** (Bangkok, December 2020)

UNFCCC

- General consideration of the use of AI in relation to climate action is being explored in the context of the [UNFCCC's Resilience Frontiers initiative](#) to further the exploration of [frontier issues](#), as launched by the United Nations Chief Executives Board for Coordination.

UNFPA

- Since 2018, **GRID3 (Geo-Referenced Infrastructure and Demographic Data for Development)** works with countries to generate, validate, and use geospatial data on population, settlements, infrastructure, and subnational boundaries in regions where an updated snapshot of populations and population distribution is needed and/or significant migration has occurred.
- **"Testing ECHO: Amplifying citizens' voices for the SDGs"** is an initiative led by UNFPA's Colombia Country Office, which is developing a tool powered by AI to promote citizens' participatory planning and awareness about the SDGs through real-time guided public discussion.

UN Global Pulse

- **Mapping the landscape of artificial intelligence applications against COVID-19³¹** - UN Global Pulse worked with researchers from the World Health Organization (WHO) and the MILA- Quebec AI Institute to provide the first comprehensive map of the landscape of AI applications that are being built to tackle the COVID-19 pandemic from March 2020.
- **On the value of ship traffic data for epidemic modeling of diseases**-UN Global Pulse set out to explore whether insights from Automated Identification Systems (AIS) data can be included in epidemic modelling of diseases, including COVID-19, to inform more efficient and timely operational responses.
- **Using AI based speech-to-text tech for epidemic intelligence and COVID19 response** UN Global Pulse is working with the WHO to explore the use of data from radio talk shows to **signal early warnings of health risks and health-related matters**. The research showed promising results for monitoring the ongoing health emergency, its social and economic impacts, and the spread of endemics.
- **Pulse Satellite: A collaboration tool using human-AI interaction to analyze satellite imagery** UN Global Pulse has been working with UNOSAT to build a software tool that leverages AI to identify map structures in refugee settlements, a roof density detection model and flood mapping applications.
- **Using artificial intelligence to model displacement in Somalia**. UN Global Pulse with the United Nations High Commissioner for Refugees (UNHCR) developed a dashboard early warning system that displays monthly predicted IDP arrivals for regions in Somalia.
- **Understanding population movement after natural disasters**. Work from Pulse Lab Jakarta and the International Organization for Migration (IOM) using data from an Indonesian telecom provider to gather insights on internal displacement of people affected by a tsunami and subsequent landslides in the Central Sulawesi province of Indonesia.
- **Deploying AI social listening tools for real-time information to UN country teams**. Partnerships to provide UN personnel with access to AI based real-time event detection tools.
- **Supporting development of ethical AI frameworks in the Global South**. AI ethics and a focus on human rights is part of a **blueprint for the development of national AI strategies** in African countries developed by the *Smart Africa Alliance*, a collaboration of 30 Africa countries, the UNGP, and the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ).

³¹ "Mapping the landscape of artificial intelligence applications against COVID-19." Bullock, et al. Journal of Artificial Intelligence Research 69 (2020): 807-845.
 "Artificial intelligence cooperation to support the global response to COVID-19." Luengo-Oroz, et al. Nature Machine Intelligence 2.6 (2020): 295-297.

- AI ethics and efforts to increase access to data in a privacy preserving manner is core to the development of **national AI strategies in Uganda and in Ghana** where UNGP is supporting representatives from the Ghanaian and Ugandan government in the development of said strategies in collaboration with GIZ.
- **Due diligence and risk assessments tools** developed by UNGP as data privacy, ethics and data protection compliance mechanisms are designed to help identify and minimize the risks of harms and maximize the positive impacts of data innovation projects.
- A **human rights due diligence tool** is in development led by UNHCR as a mechanism to assess prospective technology partners with regards to compliance with human rights supported by UNGP.
- The [Global Data Access Framework](#) is a multistakeholder initiative which aims to create a platform for sharing digital public goods in a manner that respects privacy to facilitate data sharing across the public and private sector
- UNGP issued [principles for data protection and privacy](#) including guidance on [responsible use of data in the context of the COVID-19 pandemic](#).

UNICEF

- UNICEF, in partnership with the Government of Finland, is developing an [AI for Children Policy Guidance](#) for governments and industry, which, if applied, will help to create environments that support the safe and beneficial use of AI systems for children's development. The guidance is being developed in a consultative process with expert workshops already held in Europe, Africa and Latin America. Children are also being consulted. The policy guidance will be launched in August 2020 and the world's first Global Forum on AI for Children will be held in March 2021.
- UNICEF, in collaboration with 17 global experts, is currently developing a **Manifesto for Good Governance of Children's Data**.
- UNICEF utilizes its own AI-driven platforms to engage with adolescents and it also brings together children, adolescents and youth to come up with innovative solutions related to issues affecting them, often using AI. For instance, Brazil, UNICEF has been developing an AI platform to engage adolescents on sensitive topics related to use of social media, such as girls' confidence and prevention of xenophobia in the context of the Venezuela migration crisis. In Tanzania, young innovators developed a creative solution for science and mathematics classes using AI.
- UNICEF partners with tech companies and non-profit organizations to **leverage technology and AI to address challenges facing children and adolescents**. For instance, in the East and Southern Africa region, UNICEF partnered with Foundation Botnar to develop a digital solution to address employability and skills for African Youth. UNICEF also partnered with Rockefeller Foundation in Uganda to develop intelligent community health systems.
- UNICEF partners with national governments to develop digital solutions and improve the access of vulnerable populations and children to public services.

For instance, in Romania, UNICEF has supported a National Digital Health Strategy, combining community-based services with more specialized services, including virtual components of health diagnostics driven by AI, online consultation, digital facts-for-life tools, among others. In Thailand, UNICEF supported the Ministry of Health in Developing an innovative surveillance and case management system to screen and track children and families at risk of violence and to organize preventative action, incorporating AI and machine learning so that detection rates can be improved as practice develops. As part of UNICEF life skills initiatives and supporting of such programmes all around the world, UNICEF also helps to provide training on skills related to digital literacy and AI.

- UNICEF develops knowledge products and provides information to governments supporting data governance and data use. For instance, in the UAE, UNICEF facilitated a diagnostic report surveying data landscape which led to an action plan to enhance data governance which will help facilitate global reporting of UAE's considerable data on children, and decision making in national programmes and policies.

UNIDO

- The ethical issues have been raised in the different discussions on the Fourth Industrial Revolution (4IR), including at the [Global Manufacturing and Industrialization Summit](#).
- **International Conference on Ensuring Industrial Safety: the Role of Governments, Regulations and Standards (Vienna, May 2019)** discussed the implications of several 4IR technologies like AI on industrial safety and security (safe production, safe data transfer, safe human-robots/machine interactions).

UNODC and UNICRI

- **UNODC's illicit crop monitoring program** is piloting the use of AI (machine learning and deep learning) for detection of illicit crops on satellite images.
- **Fourth Workshop of the Fourteenth United Nations Congress on Crime Prevention and Criminal Justice (Kyoto, April 2020)** is expected to include the issue of the ethical considerations, as well as procedural and human rights safeguards, in the use of technology, including artificial intelligence and robotics, against crime as one of the sub-topics of discussion.
- **Global Judicial Integrity Network** raises awareness about the implications of AI use in judiciaries through different events and advocacy methods.
- **The Centre for Artificial Intelligence and Robotics of UNICRI** has been working on AI since 2015, exploring the ethical, legal and social implications of advances in AI as they pertain to its mandate.
 - **UNICRI-INTERPOL annual Global Meeting on AI for law enforcement** since 2018.

- **Panel discussions on AI and Law Enforcement** at Tallinn Digital Summit in 2019 and at the 14th United Nations Congress on Crime Prevention and Criminal Justice.
- UNICRI and INTERPOL released a **Report on AI for Law Enforcement** in April 2019, which includes, inter alia, analysis of the ethical, legal and social implications.
- UNICRI and INTERPOL will explore the development of a toolkit for the responsible use of AI by law enforcement in 2020.

UNOOSA

- **World Space Forum – Access to Space 4 All (Vienna, November 2019)** contained a focus session on AI for satellite data interpretation in support of the SDGs, topics included the equitable access and use of AI.
- UNOOSA has partnered with ITU on the **AI for Good initiative** – Track 3 of the Summit held in Geneva in May 2018 discussed how satellite imagery and AI can support the SDGs. Key takeaways from Track 3 include an **assessment of some of the ethical implications of remote sensing** of earth including aspects of: privacy, accountability, and information asymmetries.
- A primary objective of the **UN SPIDER – Bonn International Conference (May 2019) on Space-based Solutions for Disaster Management in Africa** was the discussion of AI tools for disaster relief.

UNSG's High Level Panel on Digital Cooperation

- **Report of the High-level Panel on Digital Cooperation** provides recommendations on how the international community could work together to optimize the use of digital technologies and mitigate the risks. **Recommendation 3C of the Report** has direct relevance to the ethics of artificial intelligence.

UNU

- UNU Centre for Policy Research (UNU-CPR) in New York has worked on digital technology since 2013, this including contribution to the preparation of the **Secretary-General's Strategy on New Technologies** and the **report of the High-Level Panel on Digital Cooperation**. UNU-CPR also hosts the online thought leadership and engagement platform [AI & Global Governance](#)
- UNU-CPR has published a report entitled [The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI](#) in 2019, examining how the multilateral system can better understand and anticipate the risks that will come from AI convergence with cyber and biotechnologies.
- UNU Institute in Macau will be assembling a research team consisting of post-doctoral fellows and senior researchers well-known in the field of AI & ethics, focusing on the Global South. In particular, the Institute is setting up a **consortium on AI for social inclusion** to bring together experts in higher

education institutes and other experts in AI policy, governance, design and deployment.

WHO

- WHO has established an expert group to develop a **Guidance Document on Ethics and Governance of Artificial Intelligence for Health**.

WIPO

- WIPO has started an [open process](#) to discuss the **legal and policy implications of AI on IP**, with a list of the [main questions and issues](#) being developed concerning the impact of AI on IP policy.
- WIPO is also collating the **main government instruments of relevance to AI and IP** in an [AI and IP Strategy Clearing House](#).

Annex II: Past, ongoing and future initiatives related, either directly or indirectly, to the ethical, legal and social implications of AI within the UN system ³²

1. The United Nations System:

- a. [Secretary-General's Roadmap for Digital Cooperation](#) (2020)
- b. [First draft of the Recommendation on the Ethics of Artificial Intelligence](#) (2020)
- c. [Working Document: Toward a Draft Text of a Recommendation on the Ethics of Artificial Intelligence](#) (2020)
- d. [The Age of Digital Interdependence, Report of the UN Secretary-General's High-level Panel on Digital Cooperation](#) (2019)
- e. [The right to privacy in the digital age. Resolution adopted by the Human Rights Council](#) (2019)
- f. [A United Nations system-wide strategic approach and road map for supporting capacity development on artificial intelligence](#) (2019)
- g. [Preliminary Study on the Ethics of Artificial Intelligence](#) (2019)
- h. [Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective](#) (2019)
- i. [Final Report on the International Conference on Artificial Intelligence and Education. Planning Education in the AI Era: Lead the Leap](#) (2019)
- j. [Beijing Consensus on Artificial Intelligence and Education](#) (2019)
- k. [I'd blush if I could: closing gender divides in digital skills through education](#)
- l. [Report of COMEST on Robotics Ethics](#) (2017)[AI for Good Global Summit Report](#), ITU (2017)

2. Other international organizations:

- a. Council of Europe:
 - i. [Declaration Decl\(13/02/2019\)1 on the manipulative capabilities of algorithmic processes](#) (2019)
 - ii. [Recommendation on Artificial Intelligence and Human Rights "Unboxing artificial intelligence: 10 steps to protect human rights"](#) (2019)
 - iii. [European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment](#), Council of Europe (2018)
 - iv. [Addressing the impacts of Algorithms on Human Rights: Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems](#) (2018)

³² This list may not be comprehensive or to date. Additional resources may be suggested to the DESA secretariat for inclusion in future updated versions.

- v. [Recommendation n°2102\(2017\) about Technological convergence, artificial intelligence and human rights](#) (2017)
 - b. EU:
 - i. [White Paper on Artificial Intelligence – A European approach to excellence and trust](#), European Commission (2020)
 - ii. [EU guidelines on ethics in artificial intelligence: Context and implementation, European Parliamentary Research Service](#) (2019).
 - iii. [Statement on Artificial Intelligence, Robotics and “Autonomous” Systems](#) (incl. Ethical principles and democratic prerequisites), European Group on Ethics in Science and New Technologies, EGE (2018)
 - iv. [Ethics Guidelines for Trustworthy AI: Working Document for stakeholders’ consultation](#), European Commission’s High-Level Expert Group on AI (2018)
 - v. [Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Artificial Intelligence for Europe, COM \(2018\) 237 final](#).
 - c. OECD:
 - i. [Recommendation of the Council on Artificial Intelligence](#) (2019)
 - ii. [Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD](#) (AIGO), (2019)
 - iii. [OECD AI Policy Observatory](#)
 - iv. [OECD Network of Experts on AI](#)
 - v. [AI Wonk blog](#)
 - vi. [Repository of national AI strategies and policies](#)
 - vii. [Artificial Intelligence in Society Book](#) (2019)
 - viii. [OECD Digital Economy Outlook 2020 \(Chapter 11\)](#) (2020)
 - ix. [OECD AI Principles](#) (2019)
 - x. [OECD.AI programme on Work, Innovation, Productivity and Skills](#)
3. Member States sources:
- a. Australia
 - i. [Australia’s Ethics Framework](#), Department of Industry, Science, Energy and Resources (2019)
 - b. Canada:
 - i. A set of [guiding principles to ensure effective and ethical AI](#) (2019)
 - ii. [Canada’s Directive on Automated Decision-Making](#) (2019)
 - c. China:
 - d. [Governance Principles for the New Generation](#)
 - e. [Artificial Intelligence--Developing Responsible Artificial Intelligence \(2019\)](#)
 - f. France: [Strategy for a Meaningful Artificial Intelligence](#) (2018)
 - g. Germany:
 - i. [Opinion of the Data Ethics Commission](#) (2019)

- ii. [Automated and Connected Driving](#), BMVI Ethics Commission report (2017)
- h. Japan:
 - i. [Social Principles of Human-Centric AI](#) (2019)
 - ii. [The Japanese Society for Artificial Intelligence Ethical Guidelines](#), JSAI (2017)
 - iii. [AI R&D Principles](#), Ministry of Internal Affairs and Communications (MIC) (2017)
- i. New Zealand:
 - i. [Government Use of Artificial Intelligence in New Zealand](#) (2018)
- j. Singapore:
 - i. [Model Artificial Intelligence Governance Framework: Second Edition](#) (2020)
 - ii. [Compendium of Use Cases: Practical Illustrations of the Model AI Governance Framework](#) (2020)
 - iii. [Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organizations](#) (2020)
- k. Sweden:
 - i. [Artificial Intelligence in Swedish Business and Society](#) (2018)
- l. United Kingdom:
 - i. [Code of conduct for data-driven health and care technology](#), Department of Health and Social Care (2019)
- 4. Declarations:
 - a. [Montreal Declaration for a Responsible Development of AI](#), University of Montreal (2018)
 - b. [Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems](#), Amnesty International and Access Now (2018)
 - c. [Declaration of the Future of Life Institute on the Asilomar AI Principles](#), Future of Life Institute (2017)
- 5. Other sources:
 - a. [Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI](#), Berkman Klein Center (2020)
 - b. [Artificial Intelligence: Consumer Experiences in New Technology](#), Consumers International (2019)
 - c. [Global Technology Governance: A Multistakeholder Approach](#), World Economic Forum (2019)
 - d. [Ethically Aligned Design, First Edition: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems](#), IEEE, 2019.
 - e. [G20 Ministerial Statement on Trade and Digital Economy and G20 AI Principles](#), (2019)
 - f. [Indigenous AI](#) (2019)

- g. [SAP's Guiding Principles for Artificial Intelligence](#), SAP (2018)
- h. [Principles for AI Ethics](#), SAMSUNG (2018)
- i. [Sony Group AI Ethics Guidelines](#), Sony (2018)
- j. [Harmonious Artificial Intelligence Principles](#), HAIP (2018)
- k. [Universal Guidelines for Artificial Intelligence](#), The Public Voice (2018)
- l. [OpenAI Charter](#), OpenAI (2018)
- m. [AI at Google: Our Principles](#), Google (2018)
- n. [Microsoft AI Principles](#), Microsoft (2018)
- o. [Principles for Trust and Transparency](#), IBM (2018)
- p. [G7 Common Vision for the Future of AI](#), (2018)
- q. [Developing AI for Business with Five Core Principles](#), Sage (2017)
- r. [Principles for Algorithmic Transparency and Accountability by ACM](#), USACM (2017)
- s. [Top 10 Principles for Ethical Artificial Intelligence](#), UNI Global Union (2017)
- t. [DeepMind Ethics & Society Principles](#), DeepMind (2017)
- u. [AI Policy Principles](#), ITI (2017)
- v. [Tenets of Partnership on AI](#), PAI (2016)
- w. [Internet Governance and Regulations in Latin America](#) (2018) Chapter "Regulatory Perspective of Artificial Intelligence"
- x. Virtual South School on Internet Governance 2020 – "[Ethics and Artificial Intelligence](#)" and "[Artificial Intelligence, algorithms and daily life](#)" <https://youtu.be/8HkzGBKIEpY>

Annex III: Summary Table of Ethical Principles

<i>Principle</i>		<i>Key message</i>	<i>Sources</i>
<i>Potentially relevant principles</i>			
1	Human rights	AI should be developed and implemented in accordance with international human rights standards.	Preliminary Study on the Ethics of Artificial Intelligence (COMEST 2019)
	Principle of human rights	All artificial intelligence-related capacity-building programming by United Nations entities should respect the principles of human rights, thereby helping to ensure that a human rights-based approach should be mainstreamed into the approach to artificial intelligence adopted by Member States	A United Nations system-wide strategic approach and road map for supporting capacity development on artificial intelligence (CEB 2019)
	Human dignity	Dignity is inherent to human beings, not to machines or robots. Therefore, robots and humans are not to be confused even if an android robot has the seductive appearance of a human, or if a powerful cognitive robot has learning capacity that exceeds individual human cognition. Robots are not humans – they are the result of human creativity and they still need a technical support system and maintenance to be effective and efficient tools or mediators.	Report of COMEST on Robotics Ethics (COMEST 2017)
	Rights-based	[Internet] rooted in the Universal Declaration of Human Rights and its associated Covenants.	UNESCO General Conference 2015 decision on the Internet Universality
	Principle of respect of fundamental rights	Ensuring that the design and implementation of AI tools and services are compatible with fundamental rights.	CoE Ethical Charter 2018
	Human-centered values and fairness	i. AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality,	G20 AI Principles 2019 = OECD AI Principles 2019

		diversity, fairness, social justice, and internationally recognized labor rights. ii. To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.	
	Human rights	Ensure autonomous and intelligent systems do not infringe on internationally recognized human rights.	IEEE's Ethically Aligned Design 2019
	Secure a just transition and ensure support for fundamental freedoms and rights	As AI systems develop and augmented realities are formed, workers and work tasks will be displaced. It is vital that policies are put in place that ensure a just transition to the digital reality, including specific governmental measures to help displaced workers find new employment.	UNI Global Union 2017
2	Inclusiveness	AI should be inclusive, aiming to avoid bias and allowing for diversity and avoiding a new digital divide.	COMEST 2019
	Accessibility	[Internet] accessible to all, in both infrastructure and content.	UNESCO General Conference 2015 decision on the Internet Universality
	Diverse perspectives on the benefits and risks of AI technologies	Artificial intelligence-related capacity-building programming should gather diverse perspectives on the benefits and risks of artificial intelligence technologies and take into consideration the needs of all people, including those at risk of being left behind, especially those who are marginalized and vulnerable. People and particularly those farthest behind, including women and girls, should be at the centre of all artificial intelligence-related capacity-building programming and decision-making processes.	CEB 2019
	"Whole-of-government" and "whole-of-society" approach	Artificial intelligence-related capacity-building programming should strive to foster a "whole-of-government" and a "whole-of-society" approach, in particular in taking into account the bottom billion.	CEB 2019
	Multi-stakeholder partnerships	Artificial intelligence-related capacity-building programming should make efforts to strengthen multi-stakeholder partnerships, especially between Governments, private	CEB 2019

		sector, international organizations, civil society and academia.	
	Cooperation and synergy	All artificial intelligence-related programming by United Nations entities should actively seek cooperation and synergy with complementary developmental programmes that deliver other key elements in order to reach common goals.	CEB 2019
	Diversity inclusion principle	The development and use of AI systems must be compatible with maintaining social and cultural diversity and must not restrict the scope of lifestyle choices or personal experiences	Montreal Declaration 2018
	Inclusive growth, sustainable development and well-being	Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.	G20 AI Principles 2019 = OECD AI Principles 2019
	Share the benefits of AI systems	The economic prosperity created by AI should be distributed broadly and equally, to benefit all of humanity. Global as well as national policies aimed at bridging the economic, technological and social digital divide are therefore necessary.	UNI Global Union 2017
	Fairness and non-discrimination	With concerns about AI bias already impacting individuals globally, Fairness and Non-discrimination principles call for AI systems to be designed and used to maximize fairness and promote inclusivity. Fairness and Non-discrimination principles are present in 100% of documents in the dataset.	Berkman Klein Center 2020
3	Flourishing	AI should be developed to enhance the quality of life.	COMEST 2019
	Balancing economic, social and environmental goals	Artificial intelligence-related capacity-building programming should balance economic, social and environmental goals: reducing inequalities and ensuring equal access to opportunities, promoting productive transformation of the economy and protecting the natural environment. Such a process generates social justice within and between generations,	CEB 2019

	sustainable development, peace, and prosperity.	
Value of beneficence	Robots are useful for facilitating better safety, efficiency, and performance in many human tasks that are physically hard. Industrial robots, disaster robots, and mining robots can be used to replace human beings in dangerous environments. However, the beneficence of robots is subject to further discussion and reflection when they are designed to interact in a social context, such as in education, health care or surveillance/policing by the State.	COMEST 2017
Well-being principle	The development and use of AI systems must permit the growth of the well-being of all sentient beings	Montreal Declaration 2018
Prioritizing well-being	Prioritize metrics of well-being in the design and use of AIs because traditional metrics of prosperity do not take into account the full effect of AI systems technologies on human well-being	IEEE's Ethically Aligned Design 2019
Promotion of human values	Human Values principles state that the ends to which AI is devoted, and the means by which it is implemented, should correspond with our core values and generally promote humanity's well-being. Promotion of Human Values principles are present in 69% of documents in the dataset.	Berkman Klein Center 2020
Beneficence	While promoting good is often mentioned, it is rarely defined, though notable exceptions mention the <i>augmentation of human senses, the promotion of human well-being and flourishing, peace and happiness</i> , the creation of socio-economic opportunities, and <i>economic prosperity</i> . Similar uncertainty concerns the actors that should benefit from AI: private sector issuers tend to highlight the benefit of AI for customers, though overall many sources require AI to be shared and to benefit everyone "humanity", both of the above, "society", "as many people as possible", "all sentient creatures", the "planet" and the environment.	The global landscape of AI ethics guidelines, Nature 2019
AI must serve people and planet	Codes of ethics for the development, application and use of AI are needed so that throughout their entire operational process, AI systems remain compatible and increase the principles of human dignity, integrity, freedom,	UNI Global Union 2017

		privacy, and cultural and gender diversity, as well as fundamental human rights.	
	Well-being	Als should be used to support prosperity, health, democratic civic processes, personal freedom, goodwill, environmental sustainability, and the protection of children, people with disabilities, displaced people and other vulnerable populations.	WEF Principles Development Tool 2020
4	Autonomy	AI should respect human autonomy by requiring human control at all times. <i>NB: need to take into account situations where human control could be detrimental.</i>	COMEST 2019
	Value of autonomy	The recognition of human dignity implies that the value of autonomy does not solely concern the respect of individual autonomy, which can go as far as to refuse to be under the charge of a robot. The value of autonomy also expresses the recognition of the interdependency of relationship between humans, between humans and animals, and between humans and the environment. To what extent social robots will enrich our relationships, or reduce and standardize them? This needs to be scientifically evaluated in medical and educational practices where robots can be used, especially when vulnerable groups such as children and elderly persons are concerned. The extensive use of robots can accentuate in certain societies the rupture of social bonds. Interdependency implies that robots are part of our technical creations (part of the technocosm that we construct) and they also have environmental impacts (e-waste, energy consumption and CO2 emissions, ecological footprint) that must be considered and evaluated in the balance of benefit and risk.	COMEST 2017
	Principle “under user control”	Precluding a prescriptive approach and ensuring that users are informed actors and in control of their choices.	CoE Ethical Charter
	Respect for autonomy principle	AI systems must be developed and used while respecting people’s autonomy, and with the goal of increasing people’s control over their lives and their surroundings.	Montreal Declaration 2018
	Adopt a human-in-command approach	The development of AI must be responsible, safe and useful, where machines maintain the legal status of tools, and legal persons retain	UNI Global Union 2017

		control over, and responsibility for, these machines at all times.	
	Human control of technology	The principles under this theme require that important decisions remain subject to human review. Human Control of Technology Principles are present in 69% of documents in the dataset.	Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI (Berkman Klein Center 2020)
5	Explainability	AI should be explainable, able to provide insight into its functioning.	COMEST 2019
	Transparency and explainability	AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art: <ul style="list-style-type: none"> i. to foster a general understanding of AI systems ii. to make stakeholders aware of their interactions with AI systems, including in the workplace iii. to enable those affected by an AI system to understand the outcome iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision. 	G20 AI Principles 2019 = OECD AI Principles 2019
	Transparency and explainability	Principles under this theme articulate requirements that AI systems be designed and implemented to allow for oversight, including through translation of their operations into intelligible outputs and the provision of information about where, when, and how they are being used. Transparency and Explainability principles are present in 94% of documents in the dataset.	Berkman Klein Center 2020
	Comprehension	The reasons for any AI decisions and actions should be understood well enough for humans to control AIs for consistency with ethical principles, and to make human accountability possible.	WEF Principles Development Tool 2020

6	Transparency	The data used to train AI systems should be transparent.	COMEST 2019
	Openness	Open, in the way that Internet protocols are developed, applications are designed, and services are made available to their users.	UNESCO General Conference 2015 decision on the Internet Universality
	Principle of transparency, impartiality and fairness	Making data processing methods accessible and understandable, authorising external audits	CoE Ethical Charter
	Transparency	Ensure autonomous and intelligent systems operate in a transparent manner.	IEEE's Ethically Aligned Design 2019
	AI systems must be transparent	Workers should have the right to demand transparency in the decisions and outcomes of AI systems, as well as their underlying algorithms. They must also be consulted on AI systems implementation, development and deployment.	UNI Global Union 2017
	Transparency	Featured in 73 of our 84 sources, transparency is the most prevalent principle in the current literature. Thematic analysis reveals significant variation in relation to the interpretation, justification, domain of application and mode of achievement. References to transparency comprise efforts to <i>increase explainability, interpretability or other acts of communication and disclosure</i> . Principal domains of application include data use, human–AI interaction, automated decisions and the purpose of data use or application of AI systems. Primarily, transparency is presented as a way to minimize harm and improve AI, though some sources underline its benefit for legal reasons or to foster trust. A few sources also link transparency to dialogue, participation and the principles of democracy.	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019
7	Awareness and literacy	Algorithm awareness and a basic understanding of the workings of AI are needed to empower citizens.	COMEST 2019
	AIS technology misuse and awareness of it	Minimize the risks of misuse of AIS technology	IEEE's Ethically Aligned Design 2019

8	Responsibility	Developers and companies should take into consideration ethics when developing autonomous intelligent system.	COMEST 2019
	Principle responsibility of	Deterministic robots, and even sophisticated cognitive robots, cannot take any ethical responsibility, which lies with the designer, manufacturer, seller, user, and the State. Therefore, human beings should always be in the loop and find ways to control robots by different means (e.g. traceability, off switch, etc.) in order to maintain human moral and legal responsibility.	COMEST 2017
	Professional responsibility	These principles recognize the vital role that individuals involved in the development and deployment of AI systems play in the systems' impacts and call on their professionalism and integrity in ensuring that the appropriate stakeholders are consulted and long-term effects are planned for. Professional Responsibility principles are present in 78% of documents in the dataset.	Berkman Klein Center 2020
9	Accountability <i>(often cited in combination with responsibility or used interchangeably)</i>	Arrangements should be developed that will make possible to attribute accountability for AI-driven decisions and the behavior of AI systems.	COMEST 2019
	Accountability	AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.	G20 AI Principles 2019 = OECD AI Principles 2019
	Accountability	Ensure that designers and operators of AISs are responsible and accountable.	IEEE's Ethically Aligned Design 2019
	Ban the attribution of responsibility to robots	Robots should be designed and operated as far as is practicable to comply with existing laws, and fundamental rights and freedoms, including privacy.	UNI Global Union 2017
	Accountability	This theme includes principles concerning the importance of mechanisms to ensure that accountability for the impacts of AI systems is appropriately distributed, and that adequate remedies are provided. Accountability principles are present in 97% of documents in the dataset.	Berkman Klein Center 2020

	Accountability	The responsibility for an AI's decisions and actions should never be delegated to the AI. People should take responsibility for following ethical principles when working with AI and be held accountable when AIs break ethical principles and voluntary obligations.	WEF Principles Development Tool 2020
10	Democracy	AI should be developed, implemented and used in line with democratic principles.	COMEST 2019
	Democratic participation principle	AI systems must meet intelligibility, justifiability, and accessibility criteria, and must be subjected to democratic scrutiny, debate, and control.	Montreal Declaration 2018
11	Good governance	Governments should provide regular reports about their use of AI in policing, intelligence and security.	COMEST 2019
	Multi-stakeholder governance	Building on the successful partnerships that have evolved since WSIS between governments, the private sector, the technical and professional community, and civil society to foster the Internet's growth and use for peace, prosperity, social equality and sustainable development.	UNESCO General Conference 2015 decision on the Internet Universality
	Establish global governance mechanism	Establish multi-stakeholder Decent Work and Ethical AI governance bodies on global and regional levels. The bodies should include AI designers, manufacturers, owners, developers, researchers, employers, lawyers, civil society organizations and trade unions.	UNI Global Union 2017
12	Sustainability	<ul style="list-style-type: none"> i. For all AI applications, the potential benefits need to be balanced against the environmental impact of the entire AI and IT production cycle. ii. AI should be developed in a sustainable manner taking into account the entire AI and IT production cycle. iii. AI can be used for environmental monitoring and risk management, and to prevent and mitigate environmental crises. 	COMEST 2019
	Sustainable development principle	The development and use of AI systems must be carried out to ensure a strong environmental sustainability of the planet.	Montreal Declaration 2018
	Sustainability	To the extent that is referenced, sustainability calls for development and deployment of AI to consider protecting the environment, improving the planet's ecosystem and	The global landscape of AI ethics guidelines,

		biodiversity, contributing to fairer and more equal societies and promoting peace. Ideally, AI creates sustainable systems that process data sustainably and whose insights remain valid over time.	Nature Machine Intelligence 2019
<i>Other relevant principles or sub-principles</i>			
13	Safety and security	These principles express requirements that AI systems be safe, performing as intended, and also secure, resistant to being compromised by unauthorized parties. Safety and Security principles are present in 81% of documents in the dataset.	Berkman Klein Center 2020
	Do no harm principle	Board members emphasized the importance of incorporating the “do no harm” principle at the outset when designing solutions.	Ethics of AI Context from CEB and HLCP 2020
	‘Do not harm’ principle	‘Do not harm’ principle is a red line for robots. As many technologies, a robot has the potentiality for ‘dual-use’. Robots are usually designed for good and useful purposes (to diminish harmfulness of work for example), to help human beings, not to harm or kill them. In this regard, Isaac Asimov’s formulation of this principle (three laws) is still accurate (see paragraph 18. If we are morally serious about this ethical principle, then we have to ask ourselves whether armed drones and autonomous weapons should be banned.	COMEST 2017
	Prudence principle	The development and use of AI systems must not contribute to lessening the responsibility of human beings when decisions must be made.	Montreal Declaration 2018
	Principle of quality and security	With regard to the processing of judicial decisions and data, using certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment	CoE Ethical Charter 2018
	Robustness, security and safety	<ul style="list-style-type: none"> i. AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. ii. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system’s 	G20 AI Principles 2019 = OECD AI Principles 2019

		<p>outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.</p> <p>AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.</p>	
	Non-maleficence	References to non-maleficence occur significantly more often than references to beneficence and encompass general calls for safety and security or state that AI should never cause foreseeable or unintentional harm. More granular considerations entail the avoidance of specific risks or potential harms—for example, intentional misuse via cyberwarfare and malicious hacking—and suggest risk-management strategies. Harm is primarily interpreted as discrimination, violation of privacy or bodily harm. Less frequent characterizations include loss of trust or skills; “radical individualism”; the risk that technological progress might outpace regulatory measures; and negative impacts on long-term social well-being, infrastructure, or psychological, emotional or economic aspects.	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019
	Safety	Deliberate or inadvertent harm caused by AIs should be prohibited, prevented and stopped.	WEF Principles Development Tool 2020
14	Gender	Gender bias should be avoided in the development of algorithms, in the datasets used for their training, and in their use in decision-making.	COMEST 2019
		All artificial intelligence-related capacity-building programming by United Nations entities should be gender transformative. Gender and age transformative approaches need to be embedded in all artificial intelligence-related capacity-building programming and decision-making processes. The particular effects of artificial intelligence on women and girls, and on the increasing digital gender and age divide, should also be taken into account.	CEB 2019

		Specifically preventing the development or intensification of any discrimination between individuals or groups of individuals	CoE Ethical Charter 2018
		In the design and maintenance of AI and artificial systems, it is vital that the system is controlled for negative or harmful human-bias, and that any bias be it gender, race, sexual orientation or age is identified and is not propagated by the system.	UNI Global Union 2017
15	Age (young and elderly)	Young people have valid concerns relating to ethical issues of AI. As such, they should be included, in all their diversity, in all discussions on the ethical principles of AI and their concerns and considerations taken into account.	UNESCO Operational Strategy on Youth (2014-2021)
16	Privacy	Principles under this theme stand for the idea that AI systems should respect individuals' privacy, both in the use of data for the development of technological systems and by providing impacted people with agency over their data and decisions made with it. Privacy principles are present in 97% of documents in the dataset.	Berkman Klein Center 2020
	Value of privacy	Various protection schemes and regulations have been implemented in many countries to limit access to personal data in order to protect the privacy of individuals. However, the advent of Big Data changes the way data are collected and how they are processed (use of algorithm in profiling). The scale is much wider and the uses are expanding (e.g. commercial, state security and surveillance, research, etc.), and so are the forms of intrusion. Robots are devices that can collect data through sensors and that can use Big Data through deep learning. Therefore, collection and use of data need to be scrutinized in the design of robots, using an approach that balances the aim of the robot and the protection of privacy. Some data may be more sensitive than others; therefore a mix of approaches such as legislation, professional regulations, governance, public surveillance, etc. is necessary in order to maintain public trust in and good use of robots.	COMEST 2017
	Protection of privacy and intimacy principle	Privacy and intimacy must be protected from AI systems intrusion and data acquisition and archiving systems.	Montreal Declaration 2018

	Privacy	Ethical AI sees privacy both as a value to uphold and as a right to be protected. While often undefined, privacy is frequently presented in relation to <i>data protection and data security</i> . A few sources link privacy to freedom or trust.	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019
	Privacy	Als and people with AI responsibilities should protect personal and client data. Those who gather or share data with Als or from Als should seek and respect the preferences of those whom the data is about, including their preference to control the data.	WEF Principles Development Tool 2020
17	Solidarity principle	The development of AI systems must be compatible with maintaining the bonds of solidarity among people and generations.	Montreal Declaration 2018
	Solidarity	Solidarity as an AI principle should imply (1) sharing the prosperity created by AI, implementing mechanisms to redistribute the augmentation of productivity for all, and sharing the burdens, making sure that AI does not increase inequality and nobody is left behind; and (2) assessing the long-term implications before developing and deploying AI systems	"Solidarity should be a core ethical principle of AI." Nature Machine Intelligence 1.11 (2019): 494-494.
	Solidarity	Solidarity is mostly referenced in relation to the implications of AI for the labour market. Sources call for a strong social safety net. They underline the need for redistributing the benefits of AI in order not to threaten social cohesion and respecting potentially vulnerable persons and groups. Lastly, there is a warning of data collection and practices focused on individuals that may undermine solidarity in favor of "radical individualism".	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019
	Principle of solidarity and social justice	Any ethically permissible application should not increase disadvantage, discrimination or division in society. This principle is one of the two guiding principles proposed by the Nuffield Council, alongside the principle that any intervention should be consistent with the welfare of the future person. The French and German councils also emphasize the ethical concepts of non-maleficence and beneficence. In addition, the Deutscher Ethikrat recommends consideration of the ethical concepts of human dignity, protection of life and integrity, freedom, naturalness and responsibility.	Joint statement on the ethics of heritable human genome editing 2020

18	Value of justice (Equality)	<p>The value of justice is related to inequality. The extensive use of industrial robots and service robots will generate higher unemployment for certain segments of the work force. This raises fears concerning rising inequality within society if there are no ways to compensate, to provide work to people, or to organize the workplace differently. Work is still a central element of social and personal identity and recognition.</p> <p>The value of justice is also related to non-discrimination. Roboticists should be synthesized to the reproduction of gender bias and sexual stereotype in robots. The issue of discrimination and stigmatization through data mining collected by robots is not a trivial issue. Adequate measures need to be taken by States.</p>	COMEST 2017
	Justice, fairness and equity	Justice is mainly expressed in terms of fairness and of prevention, monitoring or mitigation of unwanted bias and discrimination, the latter being significantly less referenced than the first two by the private sector. Whereas some sources focus on justice as respect for diversity, inclusion and equality, others call for a possibility to appeal or challenge decisions or the right to redress and remedy. Sources also emphasize the importance of fair access to AI, data and the benefits of AI. Issuers from the public sector place particular emphasis on AI's impact on the labour market, and the need to address democratic or societal issues. Sources focusing on the risk of biases within datasets underline the importance of acquiring and processing accurate, complete and diverse data especially training data.	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019
	Equity principle	The development and use of AI systems must contribute to the creation of a just and equal society.	Montreal Declaration 2018
	Equality	AI should make only fair decisions consistent with human rights.	WEF Principles Development Tool 2020
19	Holistic approach	Artificial intelligence should be addressed in an ambitious and holistic manner, promoting the use of artificial intelligence as a tool in the implementation of the Goals, while also addressing emerging ethical and human rights,	CEB 2019

		decent work, technical and socioeconomic challenges.	
20	Trust	References to trust include calls for trustworthy AI research and technology, trustworthy AI developers and organizations, trustworthy “design principles”, or underline the importance of customers’ trust. Calls for trust are proposed because a culture of trust among scientists and engineers is believed to support the achievement of other organizational goals, or because overall trust in the recommendations, judgments and uses of AI is indispensable for AI to “fulfil its world changing potential”. This last point is contradicted by one guideline explicitly warning against excessive trust in AI.	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019
21	Freedom	Whereas some sources specifically refer to the freedom of expression or informational self-determination and “privacy-protecting user controls”, others generally promote freedom, empowerment or autonomy. Some documents refer to autonomy as a positive freedom, specifically the freedom to flourish, to self-determination through democratic means, the right to establish and develop relationships with other human beings, the freedom to withdraw consent, or the freedom to use a preferred platform or technology. Other documents focus on negative freedom—for example, freedom from technological experimentation ⁹⁹ , manipulation or surveillance. Freedom and autonomy are believed to be promoted through transparency and predictable AI ⁵⁵ , by not “reducing options for and knowledge of citizens”, by actively increasing people’s knowledge about AI, giving notice and consent or, conversely, by actively refraining from collecting and spreading data in absence of informed consent.	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019
22	Dignity	While dignity remains undefined in existing guidelines, save one specification that it is a prerogative of humans but not robots, there is frequent reference to what it entails: dignity is intertwined with human rights or otherwise means avoiding harm, forced acceptance, automated classification and unknown human–AI interaction. It is argued that AI	The global landscape of AI ethics guidelines, Nature Machine Intelligence 2019

		should not diminish or destroy, but respect, preserve or even increase human dignity. Dignity is believed to be preserved if it is respected by AI developers in the first place and promoted through new legislation, through governance initiatives, or through government issued technical and methodological guidelines.	
23	Remediation	Those with AI responsibilities should seek to be educated by people affected by their AIs. Workers, customers and others affected should have fair means to seek assistance or redress should AI endanger their livelihood, reputation or physical well-being.	WEF Principles Development Tool 2020
24	Addressing Bias	Given the attention currently being paid to the importance of training data used in the development of any AI system that relies on algorithmic data processing, it is crucial to ensure that these considerations are addressed in practice. As bias is hard or impossible to remove, the emphasis should be put on awareness (revealing) and auditing biases. This could be done by 1) establishing internal training programs for staff to deepen skills for ethical reflection and recognition of biases, and 2) putting in place an external, neutral entity, such as “a model testing institute” to repeatedly audit AI systems	
25	Ethics in engineering education	Education for ethical considerations and guidelines is often insufficient in the technical disciplines and throughout work-life. This needs to be addressed through changes in educational goals and priorities for technical subjects in the tertiary education as well as through provision of relevant opportunities for lifelong learning. Engineering education itself must change and incorporate broader societal views.	

Annex IV: Artificial Intelligence Terminology

AI is a complex, cross-subject, multi-purpose and an ongoing state of art. It is useful to understand AI and related concepts through some basic AI terminologies.³³

	Key Understanding	International Organizations	Private Sectors	Academia
AI	The brain of machine	Refer computer systems can perform tasks that normally require human intelligence. (ESCAP,2018)	Enables machines to exhibit human-like cognition, can drive our cars or steal our privacy, stoke corporate productivity or empower corporate spies. It can relieve workers of repetitive or dangerous tasks or strip them of their livelihoods (Mckinsey, 2017)	Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment (Nils J. Nilsson, 2010)
Weak AI	Weak AI is known as Artificial Narrow Intelligence (ANI) or Narrow AI. It refers to AI is designed to perform one specific or small set of tasks such as facial ID recognition, voice-activated assistants, autonomous driving etc. It is notable that weak AI has possessed the ability to surpass human such chess playing.	Design to perform narrow task, e.g., play chess, facial recognition, internet search, driving car. (ESCAP,2018)		
Strong AI	Strong AI is Artificial General Intelligence (AGI), in which design to perform tasks are equivalent to human.	With cognitive capacity like human is not available yet. (ESCAP, 2018)		
Super Intelligence	Artificial Super Intelligence (ASI) indicates that AI possess all human cognitive ability and its intelligence is able to overpass human in many aspects. It is still a hypothesis that required miles effort.			

³³ The provided list is a not a fully comprehensive overview of relevant terminology.

Algorithm	Algorithm is a set of rules or instructions that require computer to follow to solve problem. Algorithm is the way to implement machine learning.			
Machine Learning	It refers to the process of machines keep improving its own performance of completing designed task from existing and new data. Machine learning is the core technology of achieving AI, that requires enormous amount of data for the computer to analyse, to learn and relearn through algorithm.		It is based on algorithms that can learn from data without relying on rules-based programming. (McKinsey,2015)	
Deep Learning	A subdivision of machine learning, is designed to mimic human brain to process, identify received information step by step. Autonomous driving, healthcare diagnose are the examples of deep learning. Deep learning requires immense amount of data to train for better result.			A form of machine learning based on layered representations of variables referred to as neural networks, has made speech-understanding practical on our phones and in our kitchens, and its algorithms can be applied widely to an array of applications that rely on pattern recognition. a class of learning procedures, has facilitated object recognition in images, video labeling, and activity recognition, and is making significant inroads into other areas of perception, such as audio, speech, and natural language processing (Stanford University, 2016)
Big Data	Big data means all available data. The development of big data provides a foundation and strong support for machine learning and deep learning.			
Data Mining	Find out useful information from a large set of data through algorithm.			
Turing Test	A test developed by Alan Turing in aiming to test whether a machine can present the ability of thinking			

	like human or possess similar intelligence as human.			
--	--	--	--	--