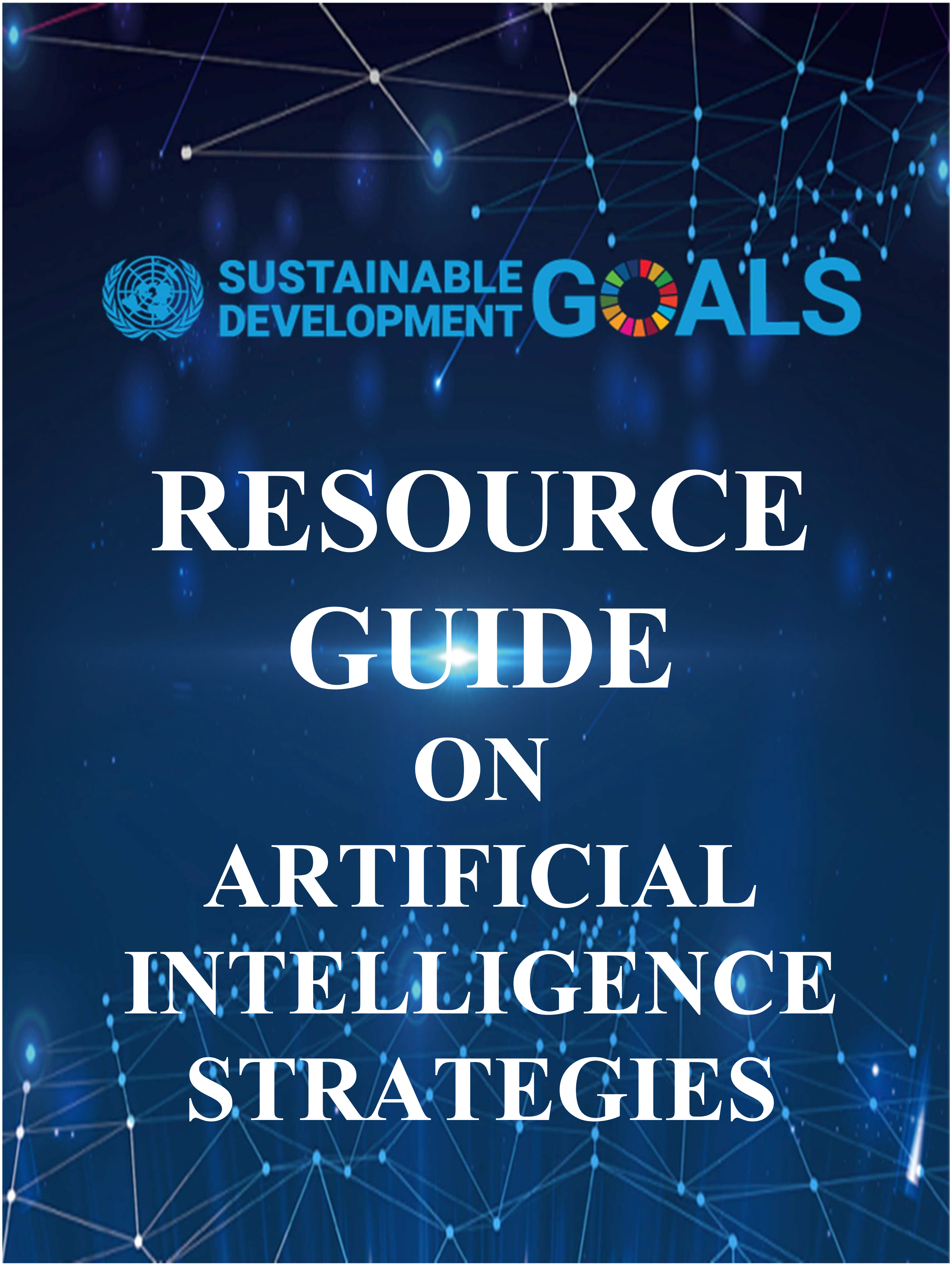


SUSTAINABLE DEVELOPMENT GOALS

RESOURCE GUIDE ON ARTIFICIAL INTELLIGENCE STRATEGIES

Resource Guide

ON

ARTIFICIAL INTELLIGENCE (AI) STRATEGIES

December 2020

(Draft for Feedback)

United Nations | Department of Economic and Social Affairs

United Nations Educational, Scientific and Cultural Organization

WFEO / FMOI

Table of Contents

Acknowledgement

The work on the Reference Guide was led by Wei Liu with input from Richard A. Roehl (DESA), under supervision by Shantanu Mukherjee (DESA).

Overall comments inputs were received from Veerle Vandeweerd, William Colglazier, Xiaolan Fu, Monika Matusiak, Naoto Kanehira, Mario Cervantes, XU Zhengzhong, and GUO Huadong.

Chapter 2 was mainly contributed by UNESCO: Dafna Feinholz, Jo Hironaka, Xianhong Hu, Melissa Tay Ru Jein, Maksim Karliuk, Sasha Rubel, Prateek Sibal, Ang Tee-Wee, Cedric Wachholz (in alphabetical order) and UNCTAD: Bob Bell Jr., Pilar Fajarnes Garces.

Chapter 3 was mainly contributed by the World Federation of Engineering Organizations (WFEO): William Kelly, GONG Ke, and LI Pan.

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

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

Chapter 1: Introduction

Artificial intelligence (AI) is becoming increasingly common in people's lives and is having a broad impact on the economy, society and environment. Meanwhile, sustainable development has been the overarching goal of the international community. Amongst numerous commitments, the United Nations called upon governments to develop national strategies for sustainable development, incorporating policy measures outlined in the 2030 Agenda for achieving the Sustainable Development Goals (SDGs). While AI technologies can be vital for breakthroughs in achieving the SDGs, they can also have unanticipated consequences, exacerbate inequalities, and constrain economic catch-up development.

The fast development of AI needs to be supported by a good overview of existing strategies to steer the growth of AI in a direction that is generally beneficial to humankind. These strategies include necessary regulatory insights and oversights for AI-based technologies to enable sustainable development. Currently, there is a lack of a coherent resource guide that can be used by policymakers and development practitioners to better understand gaps in transparency, safety, and ethical standards. At the same time, there is important knowledge and experience scattered across STI stakeholders. In this context, multi-stakeholder engagement is essential, seeing as many technology advances are initiated in the private sector and academia.

The UN Technology Facilitation Mechanism (TFM) was created by the Addis Ababa Action Agenda to support the implementation of the SDGs and launched by 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 44 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, development of practical guidance, joint 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 United Nations Interagency Task Team on Science, Technology and Innovation for the SDGs (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 fora. Member States and other key stakeholders highly appreciate this multi-stakeholder approach.

The main purpose of this Reference Guide on AI Strategies is for the TFM community to respond to the interests of- and concern related to AI related issues expressed by the stakeholders in STI Forum, 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, rather a "Resource Guide" or "Primer" mainly directed towards the diplomatic community and other policy makers who are involved in setting up global agenda on AI, with the aim of promoting more meaningful deliberations on AI related resolutions etc. 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 Roadmaps[1] focusing on one specific area, namely AI strategy development. Unlike the Guidebook on Roadmaps, this Reference Guide is not on "how to develop AI strategies" but on a collection of key references, providing a global overview of discussions on AI Ethics, technical standards, and examples of national strategies. It is planned to prepare the next version of the AI Guidebook focusing 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.[2]

[1] See: https://sustainabledevelopment.un.org/tfm#roadmaps

[2] 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 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

Chapter 2: Ethics of AI

Introduction

Artificial Intelligence (AI)[3], as a general-purpose technology, has profound implications for human beings, societies, economies and the environment. In order to unlock AI's potential to accelerate the achievement of the UN 2030 Sustainable Development Agenda, while managing risks, it is important to develop a comprehensive understanding of how societies are transformed by disruptive technologies, such as AI.

This work needs to be accompanied by an ethical reflection, as AI technologies are not value-neutral, and may influence human-technology relations in both beneficial and harmful ways. For example, AI systems may incorporate biases, due to the data on which they are based and trained, the choices made by developers while designing and training AI-algorithms on the data. Further, AI machine decisions are not always fully explainable and predictable, and thus can be difficult to understand or to redress.[4]

This chapter is divided into four sections. Section one highlights ethical implications of AI. Section two describes the call for ethically-informed approach to AI governance. Section three presents work on ethics of AI and ICTs across the UN system,

Ethical implications of AI

AI-based technologies blur the boundary between human subjects and technological objects.[5] In doing so, they not only have societal implications, which can be ethically evaluated, but they also affect the central categories of ethics: our concepts of agency[6] and responsibility, and our value frameworks.

[3] While there is no one single definition of 'artificial intelligence' (AI), this chapter tends to define 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).

[4] 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, private sector and the academia to address the black-box problem of AI, see page 79-81 of UNESCO report "Steering AI and Advanced ICTs for Knowledge Societies"

[5] For a more detailed treatment of the subject/object distinction from a philosophy of technology perspective, see UNESCO COMEST report on Robotics Ethics.

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

In terms of agency and responsibility, the increasing autonomy of AI systems raises the question, who exactly should bear ethical and/or legal responsibility for the decisions taken by AI systems. Life and death decisions to be taken by the AI system of an autonomous vehicle in the event of an accident is an example of an ethical problem of this kind.

Currently, governments are lagging in regulation and review of AI technology. Some of the world's largest corporations have attempted to set an example in the face of a regulatory gap. Amazon, for example, imposed a one-year moratorium on police access to its facial recognition technology to "give Congress enough time to put in place appropriate rules." This came at a time of widespread demonstrations recognizing racism and biased policing. Microsoft similarly urged US politicians to improve regulation on facial recognition tools and deleted its database of 10 million images that was being used to train facial recognition systems (operated by the military and police forces).

Many countries are struggling with how to regulate machine learning in decision making. New Zealand produced an Algorithm Charter for the purpose of improving government transparency and accountability in response to emerging technologies. The UK's Home Office is abandoning a decision-making algorithm that has been used to process visa applications that has come under fire for being racist.

Further, automated decision-making by AI systems may interfere with human moral agency and may have implications for our understanding of moral agency. For instance, several social media business models rely on content personalization to match an individual's prior interests, may interfere with his/her right to form opinions freely, a necessary aspect for individuals to exercise their right to freedom of expression.

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

While the questions of moral agency and moral frameworks capture broad ethical concerns raised by the development and use of AI systems, other challenges exist in the form of questions of whether humans exercise control over AI systems, the pace of cross border technology innovation and digital divide, biases embedded in algorithms, including gender biases, the protection of people's privacy and personal data, the risks of creating new forms of exclusions, the disruption of governance models, the issues of just distribution of benefits and risks, accountability, responsibility, impacts on employment and the future of work, human rights and dignity, security and risks arising out of dual use of technology.

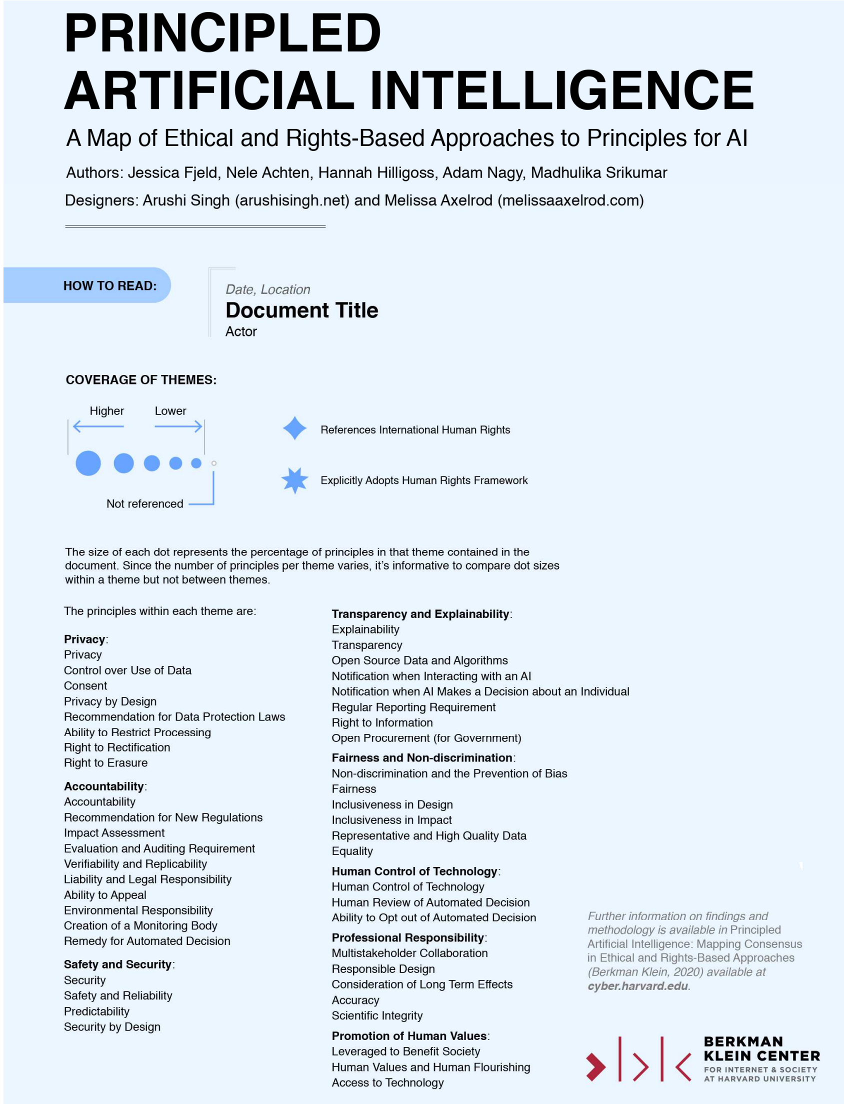
Call for an ethical approach to development and use of AI

The UN Secretary-General has underlined a need to ensure that AI becomes a force for good. Several global, regional and national initiatives on the ethical implications of AI have articulated principles and values to guide the development and use of AI systems.[7] Within the UN system, UNESCO, following a mandate from UN's 193 Member States, is developing a recommendation on the ethics of AI with an approach that is human-centered, human rights-based and respects cultural diversity,[8] after the 2018 AI conference titled "AI with human values for sustainable development".

Figures 1.1 and 1.2 provide a visualisation of different AI principles from the civil society, the private sector, governments and inter-governmental organisations. A review of different ethical principles proposed by international, regional and national initiatives is provided in Annex III.

[7] Annex III provides a links to several global, regional and national initiatives on the ethical implications of AI. Annex V provides summaries of some of these initiatives.

[8] In November 2019, the General Conference of UNESCO, at its 40th session, adopted 40 C/Resolution 37, by which it mandated the Director-General "to prepare an international standard-setting instrument on the ethics of artificial intelligence (AI) in the form of a recommendation", which is to be submitted to the General Conference at its 41st session in 2021.



PRINCIPLED ARTIFICIAL INTELLIGENCE

A Map of Ethical and Rights-Based Approaches to Principles for AI

Authors: Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, Madhulika Srikumar

Designers: Arushi Singh (arushisingh.net) and Melissa Axelrod (melissaaxelrod.com)

HOW TO READ:

Date, Location
Document Title
Actor

COVERAGE OF THEMES:

Higher ← → Lower

● ● ● • • ○

Not referenced

✦ References International Human Rights

✸ Explicitly Adopts Human Rights Framework

The size of each dot represents the percentage of principles in that theme contained in the document. Since the number of principles per theme varies, it's informative to compare dot sizes within a theme but not between themes.

The principles within each theme are:

Privacy:
Privacy
Control over Use of Data
Consent
Privacy by Design
Recommendation for Data Protection Laws
Ability to Restrict Processing
Right to Rectification
Right to Erasure

Accountability:
Accountability
Recommendation for New Regulations
Impact Assessment
Evaluation and Auditing Requirement
Verifiability and Replicability
Liability and Legal Responsibility
Ability to Appeal
Environmental Responsibility
Creation of a Monitoring Body
Remedy for Automated Decision

Safety and Security:
Security
Safety and Reliability
Predictability
Security by Design

Transparency and Explainability:
Explainability
Transparency
Open Source Data and Algorithms
Notification when Interacting with an AI
Notification when AI Makes a Decision about an Individual
Regular Reporting Requirement
Right to Information
Open Procurement (for Government)

Fairness and Non-discrimination:
Non-discrimination and the Prevention of Bias
Fairness
Inclusiveness in Design
Inclusiveness in Impact
Representative and High Quality Data
Equality

Human Control of Technology:
Human Control of Technology
Human Review of Automated Decision
Ability to Opt out of Automated Decision

Professional Responsibility:
Multistakeholder Collaboration
Responsible Design
Consideration of Long Term Effects
Accuracy
Scientific Integrity

Promotion of Human Values:
Leveraged to Benefit Society
Human Values and Human Flourishing
Access to Technology

Further information on findings and methodology is available in Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches *(Berkman Klein, 2020) available at* ***cyber.harvard.edu***.

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

Figure 1.1 Key

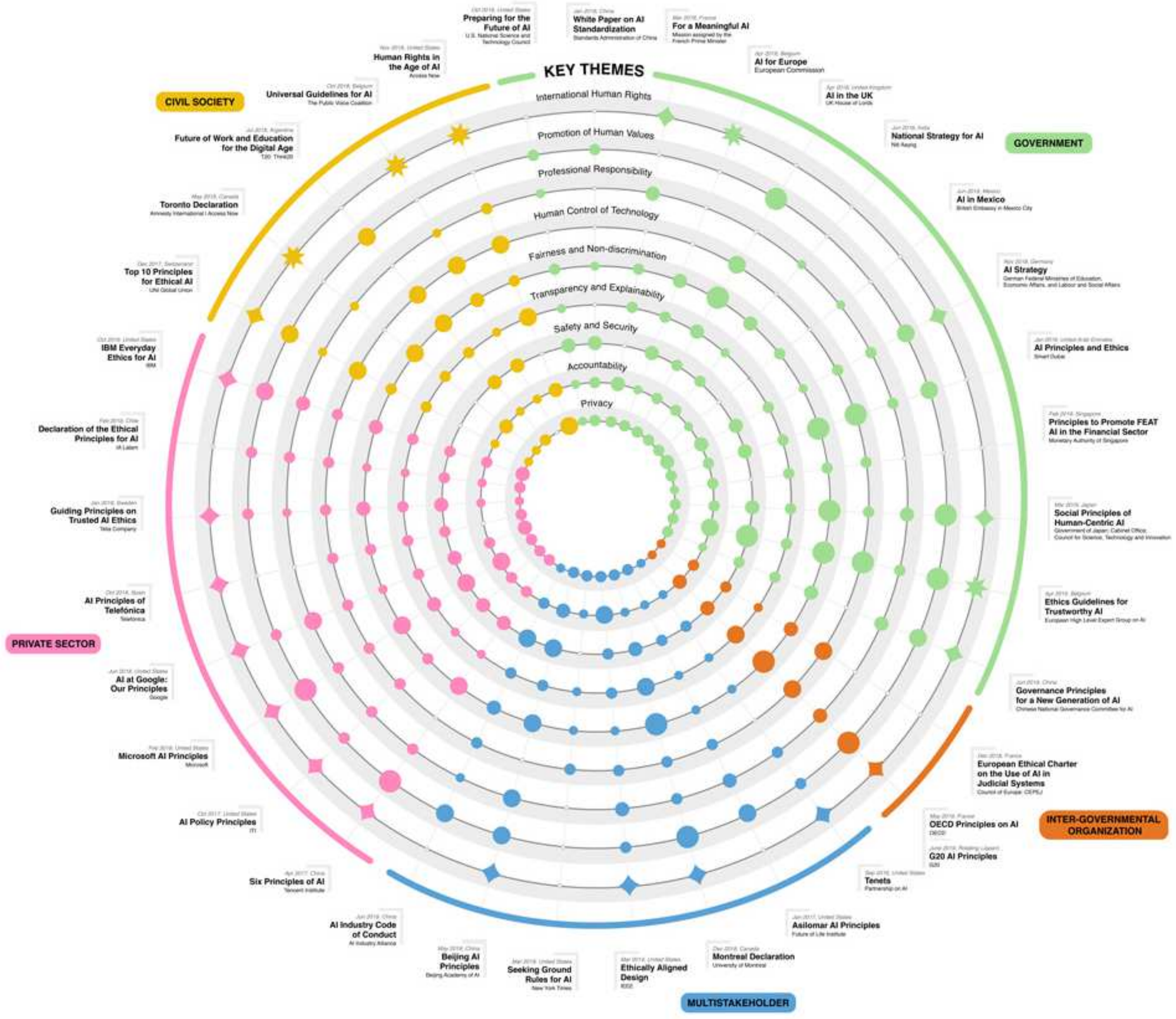

KEY THEMES
International Human Rights
Promotion of Human Values
Professional Responsibility
Human Control of Technology
Fairness and Non-discrimination
Transparency and Explainability
Safety and Security
Accountability
Privacy
CIVIL SOCIETY
GOVERNMENT
INTER-GOVERNMENTAL ORGANIZATION
MULTISTAKEHOLDER
PRIVATE SECTOR
Oct 2016, United States
Preparing for the Future of AI
U.S. National Science and Technology Council
Jan 2018, China
White Paper on AI Standardization
Standards Administration of China
Mar 2018, France
For a Meaningful AI
Mission assigned by the French Prime Minister
Apr 2018, Belgium
AI for Europe
European Commission
Apr 2018, United Kingdom
AI in the UK
UK House of Lords
Jun 2018, India
National Strategy for AI
Niti Aayog
Jun 2018, Mexico
AI in Mexico
British Embassy in Mexico City
Nov 2018, Germany
AI Strategy
German Federal Ministries of Education, Economic Affairs, and Labour and Social Affairs
Jan 2019, United Arab Emirates
AI Principles and Ethics
Smart Dubai
Feb 2019, Singapore
Principles to Promote FEAT AI in the Financial Sector
Monetary Authority of Singapore
Mar 2019, Japan
Social Principles of Human-Centric AI
Government of Japan; Cabinet Office; Council for Science, Technology and Innovation
Apr 2019, Belgium
Ethics Guidelines for Trustworthy AI
European High Level Expert Group on AI
Jun 2019, China
Governance Principles for a New Generation of AI
Chinese National Governance Committee for AI
Dec 2018, France
European Ethical Charter on the Use of AI in Judicial Systems
Council of Europe: CEPEJ
May 2019, France
OECD Principles on AI
OECD
June 2019, Rotating (Japan)
G20 AI Principles
G20
Sep 2016, United States
Tenets
Partnership on AI
Jan 2017, United States
Asilomar AI Principles
Future of Life Institute
Dec 2018, Canada
Montreal Declaration
University of Montreal
Mar 2019, United States
Ethically Aligned Design
IEEE
Mar 2019, United States
Seeking Ground Rules for AI
New York Times
May 2019, China
Beijing AI Principles
Beijing Academy of AI
Jun 2019, China
AI Industry Code of Conduct
AI Industry Alliance
Apr 2017, China
Six Principles of AI
Tencent Institute
Oct 2017, United States
AI Policy Principles
ITI
Feb 2018, United States
Microsoft AI Principles
Microsoft
Jun 2018, United States
AI at Google: Our Principles
Google
Oct 2018, Spain
AI Principles of Telefónica
Telefónica
Jan 2019, Sweden
Guiding Principles on Trusted AI Ethics
Telia Company
Feb 2019, Chile
Declaration of the Ethical Principles for AI
IA Latam
Oct 2018, United States
IBM Everyday Ethics for AI
IBM
Dec 2017, Switzerland
Top 10 Principles for Ethical AI
UNI Global Union
May 2018, Canada
Toronto Declaration
Amnesty International | Access Now
Jul 2018, Argentina
Future of Work and Education for the Digital Age
T20: Think20
Oct 2018, Belgium
Universal Guidelines for AI
The Public Voice Coalition
Nov 2018, United States
Human Rights in the Age of AI
Access Now

Figure 1.2: Compilation of Principles for Artificial Intelligence development and use (Source: Berkman Klein Center).

United Nations work on AI Ethics

A history of engagement with ethics of technology

The work on Ethics of AI builds on a long history, within the UN system, of engagement with ethical concerns related to development and use of information and communication technologies.

The 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, and religion in conformity with relevant international instruments."

In 2015, UNESCO Member States committed to promoting human rights-based ethical reflection, research and public dialogue on the implications of new and emerging technologies and their potential societal impacts with the adoption of the Internet Universality framework and the associated R.O.A.M principles.[9]

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.[10]

The draft text of the Recommendation will provide an opportunity for Member States to discuss and agree upon an initial non-exhaustive set of basic principles and recommended policy actions as ethical and human rights guardrails for the design, development and deployment of AI. It will also address the concerns of developing countries, the good 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 roadmap for the development of recommendation is presented in Annex I.

[9] An acronym for Human Rights, Openness, Accessibility to all, and Multistakeholder participation

[10] 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).

The draft text of the Recommendation will draw 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 will link 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 Councils resolution on "The right to privacy in the digital age", and others. More information on the AI ethics related initiatives of UN entities is included in Annex II.

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.[11]

[11] 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/

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

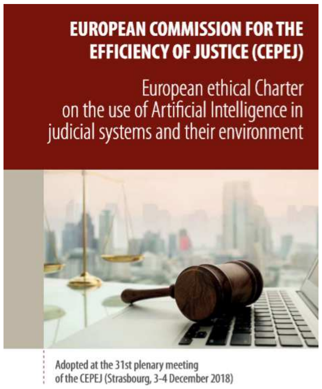
Key References

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

International Organisations

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

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



EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ)

European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment

Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 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.

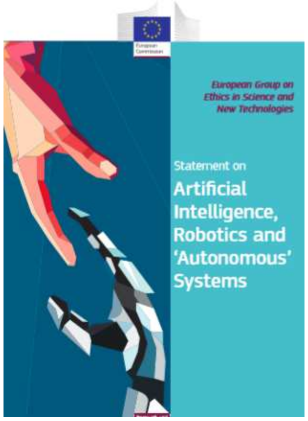
The five principles are: respect for fundamental rights; non-discrimination; quality and security (use certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment); transparency, impartiality and fairness; and "under user control"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

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

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


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

This statement 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.

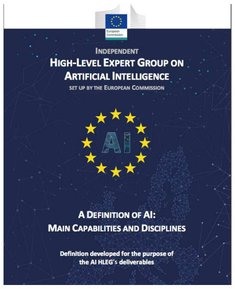
The principles within the statement are: respect for human dignity (limit to use of algorithms to affect individuals and right to know and decide whether one is interacting with a machine), autonomy (being able to set one's own standards and live according to them, being able to intervene in autonomous 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 jeopardised 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 friendlinesss).

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

Download here

3. Ethics Guidelines for Trustworthy AI (European High-Level Expert Group on 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-discrimination and fairness; societal and environmental well-being; accountability	European High-Level Expert Group on AI	Apr 2019


European Commission
Independent
High-Level Expert Group on Artificial Intelligence
set up by the European Commission
AI
A Definition of AI:
Main Capabilities and Disciplines
Definition developed for the purpose of the AI HLEG's deliverables

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 emphasises 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. G20 Ministerial Statement on Trade and Digital Economy (G20)

Key words	Developed by	Year
human-centred AI; inclusive growth, sustainable development and well-being; fairness, rule of law, human rights, and democratic values; transparency and explainability; robustness, security, safety; accountability	G20	Jun 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-centred AI to improve the work environment and quality of life, and created a future society with opportunities for everyone, including women and girls and vulnerable groups. Principles included are: inclusive growth, sustainable development and well-being; human-centred values and fairness, respecting rule of law, human rights, and democratic values; transparency and explainability; robustness, security and safety; and accountability.
Download here

G20 Ministerial Statement on Trade and Digital Economy

1. We, the G20 Trade Ministers and Digital Economy Ministers, met on 8 and 9 June 2019 in Tsukuba City, Ibaraki Prefecture, Japan, under the chairmanship of H.E. Mr. Hiroshige Seko, Minister of Economy, Trade and Industry, H.E. Mr. Masatoshi Ishida, Minister for Internal Affairs and Communications, and H.E. Mr. Taro Kono, Minister for Foreign Affairs, of the Government of Japan, to further strengthen G20 trade and digital economic policy cooperation.
2. The G20 Ministerial Meeting on Trade and Digital Economy gathered all G20 members as well as guests from Chile as 2019 APEC host economy, Egypt on behalf of AU, Estonia (for Digital Economy), Netherlands, Nigeria (for Trade), Senegal on behalf of NEPAD, Singapore, Spain, and Viet Nam. International Organizations[1]also participated in the Meeting.
3. We discussed the need to do more to achieve our common objectives for global growth. International trade and investment should continue to be important engines of growth, productivity, innovation, job creation and development.
4. Innovative digital technologies continue to bring immense economic opportunities. At the same time, they continue to create challenges.
5. This Meeting, which gathered both Trade Ministers and Digital Economy Ministers together for the first time, presented an opportunity to deepen our understanding of the interface between trade and the digital economy.
6. We discussed how we can work together toward the realization of a sustainable and innovative global society, by making full use of digital technologies, together with trade and investment, and harnessing the benefits of technological transformation and globalization, taking into consideration national needs, priorities and circumstances.

5. Charlevoix Common Vision for the Future of AI (G7)

Key words	Developed by	Year
Multistakeholder; human-centric; personal data and privacy protection; inclusivity and empowerment of women and marginalised communities; accountability, assurance, liability, security, safety; transparency	G7	Jan 2019

The Vision 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 multistakeholder, human-centric approach.

Some principles which emerge from the twelve commitments within the vision include: human-centricity and personal data and privacy protection; inclusivity and empowerment of women and marginalised communities; accountability, assurance, liability, security, safety; multistakeholder dialogue; and transparency.

Download here

6. Declaration on Ethics and Data Protection in AI (ICDPPC)

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; multistakeholder	ICDPPC	Oct 2018

Declaration on Ethics and Data Protection in Artificial Intelligence

ICDPPC

International Conference of Data Protection & Privacy Commissioners

DECLARATION ON ETHICS AND DATA PROTECTION IN ARTIFICAL INTELLIGENCE

40th International Conference of Data Protection and Privacy Commissioners

Tuesday 23rd October 2018, Brussels

AUTHORS:

- Commission Nationale de l'Informatique et des Libertés (CNIL), France
- European Data Protection Supervisor (EDPS), European Union
- Garante per la protezione dei dati personali, Italy

CO-SPONSORS:

- Agencia de Acceso a la Información Pública, Argentina
- Commission d'accès à l'information, Québec, Canada
- Datatilsynet (Data Inspectorate), Norway
- Information Commissioner's Office (ICO), United Kingdom
- Préposé fédéral à la protection des données et à la transparence, Switzerland
- Data protection Authority, Belgium
- Privacy Commissioner for Personal Data, Hong-Kong
- Data protection Commission, Ireland
- Data Protection Office, Poland
- Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), Mexico
- National Authority for Data Protection and Freedom of Information, Hungary
- Federal Commissioner for Data Protection and Freedom of Information, Germany
- Office of the Privacy Commissioner (OPC), Canada
- National Privacy Commission, Philippines

1 | Page

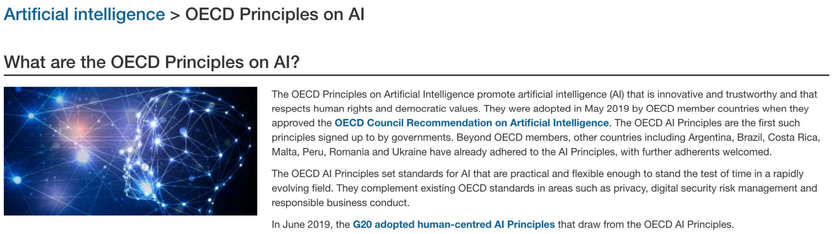
Written and sponsored by eighteen data protection and privacy commissioners across the world, the declaration fundamentally aims to preserve human rights in the development of AI. Its guiding principles include respect for human rights and fairness, constant monitoring and accountability, transparency 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

7. OECD Principles on AI (OECD)

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



Artificial intelligence > OECD Principles on AI

What are the OECD Principles on AI?

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 OECD member countries when they approved the **OECD Council Recommendation on Artificial Intelligence**. 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 already adhered to the AI Principles, with further adherents welcomed.

The OECD AI Principles set standards for AI that are practical and flexible enough to stand the test of time in a rapidly evolving field. They complement existing OECD standards in areas such as privacy, digital security risk management and responsible business conduct.

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

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.

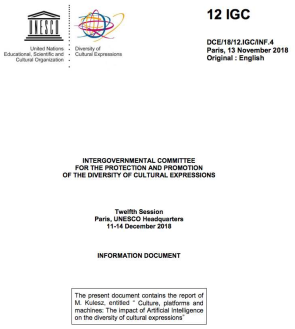
There are five principles: AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being; AI systems should be designed to respect rule of law, human rights, democratic values and diversity; there should be transparency and responsible disclosure to ensure 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 organisations and individuals developing, deploying or operating AI systems should be held accountable.

Download here

8. Culture, Platforms and Machines: the Impact of AI on the Diversity of Cultural Expressions (UNESCO)

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.



United Nations Educational, Scientific and Cultural Organization

Diversity of Cultural Expressions

12 IGC

DCE/18/12.IGC/INF.4
Paris, 13 November 2018
Original : English

INTERGOVERNMENTAL COMMITTEE
FOR THE PROTECTION AND PROMOTION
OF THE DIVERSITY OF CULTURAL EXPRESSIONS

Twelfth Session
Paris, UNESCO Headquarters
11-14 December 2018

INFORMATION DOCUMENT

The present document contains the report of M. Kulesz, entitled " Culture, platforms and machines: The impact of Artificial Intelligence on the diversity of cultural expressions"

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, emphasising that AI should not homogenise cultural expressions but rather should be used to provide better access to varied expressions and promote perspectives of traditionally marginalised groups.

Download here

9. Beijing Consensus on AI and Education (UNESCO)

Key words	Developed by	Year
humanistic; human rights; sustainable development; human-controlled; service of people and enhance human capacities; ethical, non-discriminatory, equitable, transparent and auditable manner; monitoring	UNESCO	May 2019

United Nations
Educational, Scientific and
Cultural Organization

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura

Организация
Объединенных Наций по
вопросам образования,
науки и культуры

منظمة الأمم المتحدة
للتربية والعلم والثقافة

联合国教育、
科学及文化组织

BEIJING CONSENSUS
on artificial intelligence and education

CONSENSUS DE BEIJING
sur l'intelligence artificielle et l'éducation

CONSENSO DE BEIJING
sobre la inteligencia artificial y la educación

ПЕКИНСКИЙ КОНСЕНСУС
по искусственному интеллекту и образованию

北京共识
人工智能与教育

توافق بيجين
بشأن الذكاء الاصطناعي والتعليم

Education 2030

This first-ever document 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.

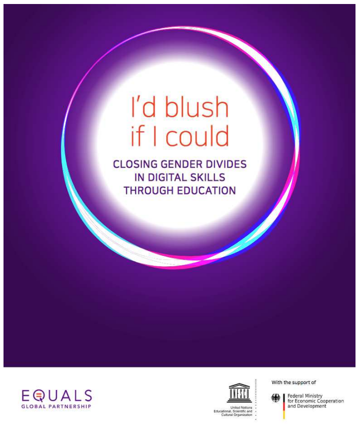
The consensus reaffirms a humanistic approach to the use of AI with a view towards protecting human rights and preparing all people with the appropriate values and skills needed for effective human–machine collaboration in life, learning and work, and for sustainable development. It advocates for human-controlled and human-centred 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

10. I'd Blush If I Could (UNESCO)

Key words	Developed by	Year
Gender equality	UNESCO	May 2019


I'd blush if I could
CLOSING GENDER DIVIDES IN DIGITAL SKILLS THROUGH EDUCATION
EQUALS GLOBAL PARTNERSHIP
With the support of
Federal Ministry for Economic Cooperation and Development

This publication 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.

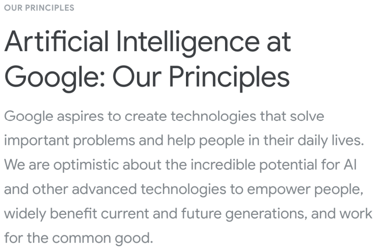
The publication provides evidence for the gender gap; for example, in numerous countries, women are 25 per cent less likely than men to know how to leverage ICT for basic purposes. It also describes several reasons why this is detrimental to women, including compromising their safety; preventing them 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; normalisation of tolerance of sexual harassment and verbal abuse; and portraying women as the voices of servility and dumb mistakes. Potential solutions suggested include including women in the process of developing AI, applying gender-sensitive approaches to digital skills development; creating new tools, processes and ensuring oversight.

Download here

Private Sector

1. AI at Google: Our Principles (Google)

Key words	**Developed by**	**Year**
socially beneficial; public availability of information; avoid bias; safety; accountable; privacy; scientific excellence	Google	Jun 2018



OUR PRINCIPLES

Artificial Intelligence at Google: Our Principles

Google aspires to create technologies that solve important problems and help people in their daily lives. We are optimistic about the incredible potential for AI and other advanced technologies to empower people, widely benefit current and future generations, and work for the common good.

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.

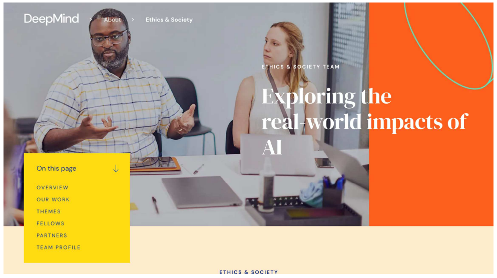
Google commits to developing AI applications based on the following objectives: be socially beneficial and facilitate the public availability of high-quality and accurate information while respecting cultural, social and legal norms; avoid creating or reinforcing unfair bias; be built and tested for 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

2. DeepMind Ethics & Society Principles (DeepMind)

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

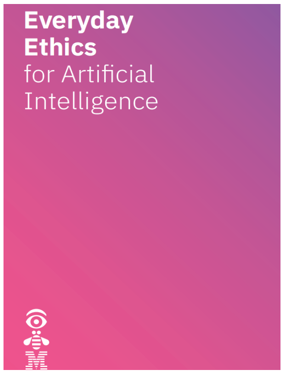


It is important to note that the following are not principles that DeepMind has committed to, but rather research areas that are being explored. These thematic areas are a creation of an independent research unit of DeepMind, a subsidiary of Google's parent company Alphabet Inc. The preamble to the principles 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; and economic impact: inclusion and equality. The last area's focus on the economic/employment impact is unique among private firms.
Download here

3. Everyday Ethics for AI (IBM)

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 (minimise 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

4. Guiding Principles for AI (SAP)

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

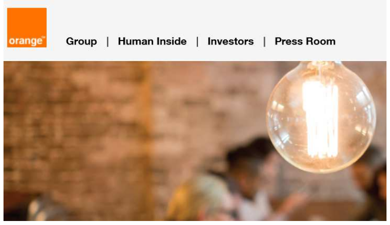
About SAP SE / SAP News Center / Technology
SAP's Guiding Principles for Artificial Intelligence
September 18, 2018 by Corinna Machmeier

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

5. Human Inside (Orange)

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

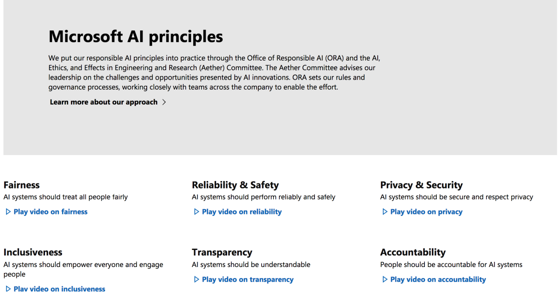


While Orange does not have a clearly defined framework of AI ethics, it emphasises the responsible use of AI, guided by its broader philosophy of "Human Inside", which it applies to all of its work. According to this framework, Orange's technical contribution is meant to be beneficial to individuals,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

6. Microsoft AI Principles (Microsoft)

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



Microsoft AI principles

We put our responsible AI principles into practice through the Office of Responsible AI (ORA) and the AI, Ethics, and Effects in Engineering and Research (Aether) Committee. The Aether Committee advises our leadership on the challenges and opportunities presented by AI innovations. ORA sets our rules and governance processes, working closely with teams across the company to enable the effort.

Learn more about our approach >

Fairness
AI systems should treat all people fairly
▷ Play video on fairness

Reliability & Safety
AI systems should perform reliably and safely
▷ Play video on reliability

Privacy & Security
AI systems should be secure and respect privacy
▷ Play video on privacy

Inclusiveness
AI systems should empower everyone and engage people
▷ Plav video on inclusiveness

Transparency
AI systems should be understandable
▷ Play video on transparency

Accountability
People should be accountable for AI systems
▷ Play video on accountability

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.

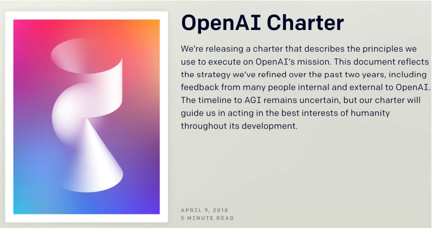
The Office of Responsible AI (ORA) and the AI, Ethics, and Effects in Engineering and Research (Aether) Committee are responsible for putting these principles into practice. Specific actions taken by Microsoft include applying these principles to their own research and work; helping other organisations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

Download here

7. OpenAI Charter (OpenAI)

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.



OpenAI Charter

We're releasing a charter that describes the principles we use to execute on OpenAI's mission. This document reflects the strategy we've refined over the past two years, including feedback from many people internal and external to OpenAI. The timeline to AGI remains uncertain, but our charter will guide us in acting in the best interests of humanity throughout its development.

APRIL 9, 2018
5 MINUTE READ

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

8. Principles for Trust and Transparency (IBM)

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

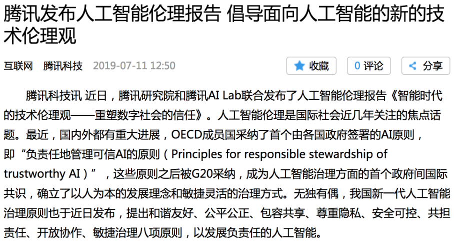
Data Responsibility

IBM's Principles for Trust and Transparency

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

9. The Ethics of Technology in the Intelligent Age - Reshaping Trust in a Digital Society (Tencent)

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



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

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。

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; and 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

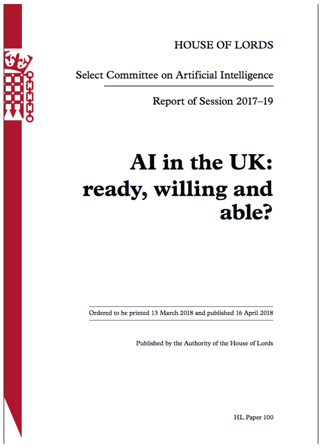
10. TrUE AI Approach (Thales Group)

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

National Governments

1. AI Code (UK House of Lords)



HOUSE OF LORDS

Select Committee on Artificial Intelligence

Report of Session 2017–19

AI in the UK: ready, willing and able?

Ordered to be printed 13 March 2018 and published 16 April 2018

Published by the Authority of the House of Lords

HL Paper 100

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

The AI code is embedded within the broader report on the economic, ethical and social implications of advances in AI for the UK. The formulation of the framework is actually recommended as part of UK's strategy for AI – judging that the UK cannot compete with the United States and China in developing AI, the report proposes that UK should instead take the lead in the ethical development and use of AI.

Considering this factor and after having consulted with organisations who have released AI ethics frameworks, the report suggests the development of a core set of widely recognised 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 organisations 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 'AII code', which could provide the basis for statutory regulation if it is deemed necessary

Download here

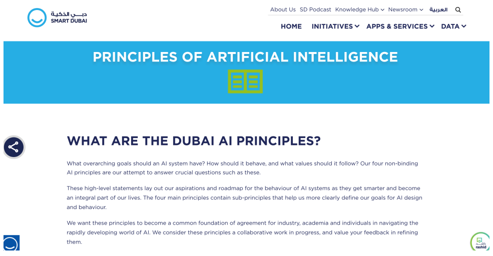
2. AI Principles and Ethics (Smart Dubai)

Key words	Developed by	Year
ethics; fair; transparent; accountable; understandable; security; safe; serve and	Smart Dubai	Jan 2019

protect humanity; beneficial to humanity; aligned with human values; inclusiveness; global governance; respect dignity and rights

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.



PRINCIPLES OF ARTIFICIAL INTELLIGENCE

WHAT ARE THE DUBAI AI PRINCIPLES?

What overarching goals should an AI system have? How should it behave, and what values should it follow? Our four non-binding AI principles are our attempt to answer crucial questions such as these.

These high-level statements lay out our aspirations and roadmap for the behaviour of AI systems as they get smarter and become an integral part of our lives. The four main principles contain sub-principles that help us more clearly define our goals for AI design and behaviour.

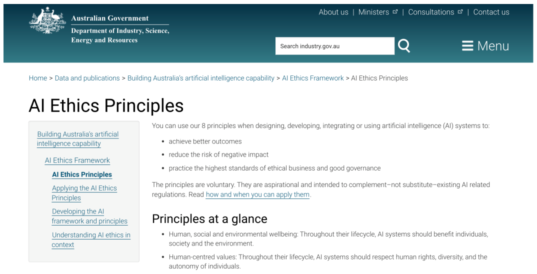
We want these principles to become a common foundation of agreement for industry, academia and individuals in navigating the rapidly developing world of AI. We consider these principles a collaborative work in progress, and value your feedback in refining them.

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

3. Australia’s Ethics Framework (Australia Department of Industry, Innovation and Science)

Key words	**Developed by**	**Year**
Human, social and environmental well-being; human-centred 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


Australian Government
Department of Industry, Science, Energy and Resources
About us | Ministers | Consultations | Contact us
Search industry.gov.au
Menu

Home > Data and publications > Building Australia's artificial intelligence capability > AI Ethics Framework > AI Ethics Principles

AI Ethics Principles

Building Australia's artificial intelligence capability
AI Ethics Framework
AI Ethics Principles
Applying the AI Ethics Principles
Developing the AI framework and principles
Understanding AI ethics in context

You can use our 8 principles when designing, developing, integrating or using artificial intelligence (AI) systems to:

- achieve better outcomes
- reduce the risk of negative impact
- practice the highest standards of ethical business and good governance

The principles are voluntary. They are aspirational and intended to complement–not substitute–existing AI related regulations. Read how and when you can apply them.

Principles at a glance

- Human, social and environmental wellbeing: Throughout their lifecycle, AI systems should benefit individuals, society and the environment.
- Human-centred values: Throughout their lifecycle, AI systems should respect human rights, diversity, and the autonomy of individuals.

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.

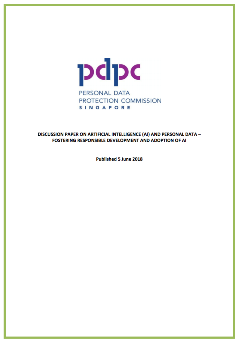
The principles are: human, social and environmental wellbeing; human-centred values (human rights, diversity, autonomy); fairness (inclusive, accessible, no unfair discrimination); privacy protection and security (privacy rights, data protection, data security); reliability and safety (operate in accordance 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

4. Discussion Paper on Artificial Intelligence (AI) and Personal Data (Personal Data Protection Commission Singapore)

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

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 is operationalises 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.



pdpc
PERSONAL DATA
PROTECTION COMMISSION
SINGAPORE

DISCUSSION PAPER ON ARTIFICIAL INTELLIGENCE (AI) AND PERSONAL DATA – FOSTERING RESPONSIBLE DEVELOPMENT AND ADOPTION OF AI

Published 5 June 2018

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).

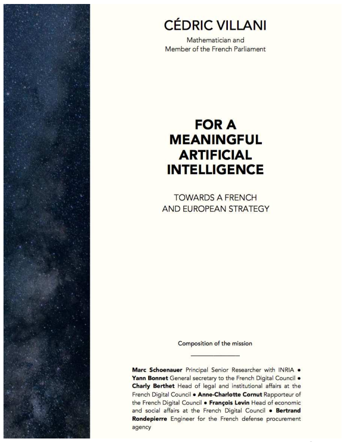
The four stages of the framework are: identifying the objectives of a governance framework; selecting appropriate organisational governance measures; considering consumer relationship management processes; and building a decision-making and risk assessment framework.

Download here

5. For a Meaningful Artificial Intelligence. Towards a French and European Strategy (Mission Villani)

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

and political debate transparency and auditability



CÉDRIC VILLANI
Mathematician and
Member of the French Parliament

FOR A
MEANINGFUL
ARTIFICIAL
INTELLIGENCE

TOWARDS A FRENCH
AND EUROPEAN STRATEGY

Composition of the mission

Marc Schoenauer Principal Senior Researcher with INRIA • **Yann Bonnet** General secretary to the French Digital Council • **Charly Berthet** Head of legal and institutional affairs at the French Digital Council • **Anne-Charlotte Cornut** Rapporteur of the French Digital Council • **François Levin** Head of economic and social affairs at the French Digital Council • **Bertrand Rondepierre** Engineer for the French defense procurement agency

The specific AI principles advocated by Mission Villani are embedded within the larger France AI Strategy which aims to foster a meaningful AI, a facet of which is ethical considerations. Ethics of AI are specifically considered in Part 5 of the strategy document, but Parts 4 and 5 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, 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 explanability, 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. Governance Principles for a New Generation of AI (Chinese National Governance Committee for 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

www.most.gov.cn

发展负责任的人工智能：新一代人工智能治理原则发布

日期：2019年06月17日　13:56　　来源：科技部

The Governance Committee for AI, linked to the Ministry of Science and Technology, outlined eight principles to "promote the 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 or English here

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

HOW CAN HUMANS
KEEP THE UPPER HAND?

The ethical matters raised by algorithms and artificial intelligence

REPORT ON THE PUBLIC DEBATE LED BY THE FRENCH DATA PROTECTION AUTHORITY (CNIL)
AS PART OF THE ETHICAL DISCUSSION ASSIGNMENT SET BY THE DIGITAL REPUBLIC BILL

DECEMBER 2017

CNIL.
COMMISSION NATIONALE INFORMATIQUE & LIBERTÉS

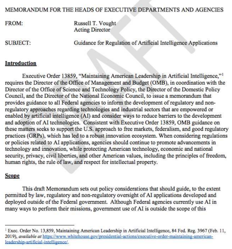
Key words	Developed by	Year
fairness; continued attention and vigilance; requirement for human intervention; intelligibility , transparency and accountability	French Data Protection Authority (CNIL)	Dec 2017

The report is the result of public debate organised 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 hybridisation of humans and machines, of which the threat to collective principles and hybridisation 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 multistakeholders human deliberation governs and guides the use of algorithms and its effects); and intelligibility, transparency and accountability.

Download here

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

Key words	Developed by	Year
public trust in AI; public participation; scientific integrity and information quality; risk assessment and management; benefits and costs; flexibility; fairness and non-discrimination; disclosure and transparency; safety and security; Interagency coordination	The White House Office of Science and Technology Policy (OSTP), United States	Jan 2020



MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought
Acting Director

SUBJECT: Guidance for Regulation of Artificial Intelligence Applications

Introduction

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 artificial intelligence (AI) and consider ways to reduce barriers to the development and adoption of AI technologies. Consistent with Executive Order 13859, OMB guidance on these matters seeks to support the U.S. approach to free markets, federalism, and good regulatory practices (GRPs), 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.

Scope

This draft Memorandum sets out policy considerations that should guide, to the extent permitted by law, regulatory and non-regulatory oversight of AI applications developed and deployed outside of the Federal government. Although Federal agencies currently use AI in many ways to perform their missions, government use of AI is outside the scope of this

[1] Exec. Order No. 13,859, Maintaining American Leadership in Artificial Intelligence, 84 Fed. Reg. 3967 (Feb. 11, 2019), *available at* https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

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 emphasise the need to encourage innovation and growth by minimising the regulatory burden, and this attention to minimised regulation is reiterated throughout the description of every principle.

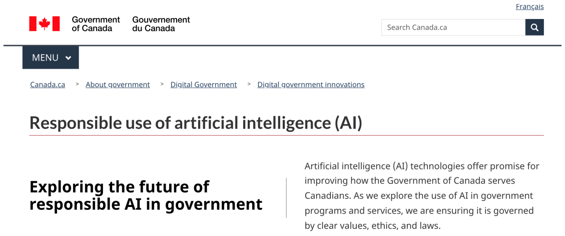
The principles are: public trust in AI (privacy, individual rights, autonomy, civil liberties); public participation and awareness; scientific integrity and information quality (quality, transparency, compliance, bias mitigation, appropriate uses, predictable, reliable and optimised outcomes); risk assessment 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

9. Responsible use of artificial intelligence (AI): Our guiding principles (Canada)

Key words	Developed by	Year
understanding and measuring impact of AI; transparent; public benefit; meaningful explanations; opportunities to review; open; protecting personal information; providing training; responsible design	Canada	Sep 2019

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



Français
Government of Canada
Gouvernement du Canada
Search Canada.ca
MENU
Canada.ca > About government > Digital Government > Digital government innovations

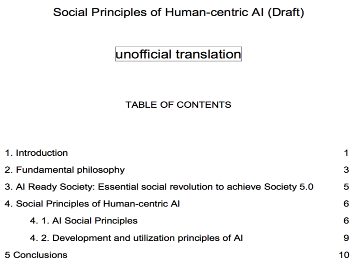
Responsible use of artificial intelligence (AI)

Exploring the future of responsible AI in government

Artificial intelligence (AI) technologies offer promise for improving how the Government of Canada serves Canadians. As we explore the use of AI in government programs and services, we are ensuring it is governed by clear values, ethics, and laws.

government. The principles are: understanding and measuring the impact of using AI by developing and sharing tools and approaches; being transparent about how and when we are using AI, starting with a clear user need and public benefit; providing meaningful explanations about AI decision making, while also offering opportunities to review results and challenge these decisions; b**e** as open as possible by sharing source code, training data, and other relevant information, all while protecting personal information, system integration, and national security and defence; and providing sufficient training so that government employees developing and using AI solutions have the responsible design, function, and implementation skills needed to make AI-based public services better.
Download here

10. Social Principles of Human-Centric AI (Government of Japan)

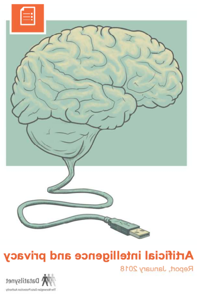
Key words	**Developed by**	**Year**
dignity; inclusion and diversity; sustainability; human-centric; human rights; expand human abilities; happiness; prevent overreliance and malicious use; human autonomy and control; user-friendliness; education; eliminate disparities; privacy; accuracy and legitimacy;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

Released by the Cabinet Office, these principles follow the publication of Japan's AI strategy. The principles, together with individual organisations' principles of AI development and utilisation, 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

11. Artificial Intelligence and Privacy (Norwegian Data Protection Authority)

Key words	Developed by	Year
privacy; data protection; fairness; purpose limitation; data minimisation; 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

Civil Society

1. Artificial Intelligence and Machine Learning: Policy Paper (Internet Society)

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

Artificial Intelligence and Machine Learning: Policy Paper

Internet Society

April 2017

Foreword

Artificial intelligence (AI) is a technology that is already impacting how users interact with, and are affected by, the Internet. In the near future, its impact is likely to only continue to grow. AI has the potential to vastly change the way that humans interact, not only with the digital world, but also with each other, through their work and through other socio-economic institutions – for better or for worse.

If we are to ensure that the impact of artificial intelligence will be positive, it will be essential that all stakeholders participate in the debates 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 its challenges can be addressed and the vast benefits the technology offers 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

2. Asilomar AI Principles (Future of Life Institute)

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

ASILOMAR AI PRINCIPLES

These principles were developed in conjunction with the 2017 Asilomar conference (videos here), through the process described here.

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 operationalised. 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 organisation.

Download here

3. Beijing AI Principles (Beijing Academy of AI)

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; optimising employment; harmony and cooperation; adaptation and moderation; subdivision and implementation; long-term planning	Beijing Academy of AI 智源动态 Beijing AI Principles 2019年5月28日	Jun 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 with no guidelines for operationalisation. Some of the principles also go beyond ethical values, into policy recommendations.

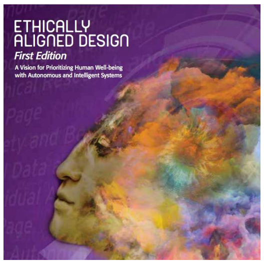
For research and development, the principles are: AI should do good (promote progress of society and human civilisation, promote sustainable development); be for humanity (conform to human values and interests such as privacy, dignity, freedom, autonomy, rights, and should not be used against humans); 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: optimising 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

4. Ethically Aligned Design (IEEE)


ETHICALLY ALIGNED DESIGN
First Edition
A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems

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.

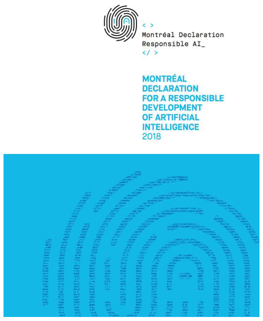
The general principles identified are: protecting and promoting human rights; increased well-being; data agency and control over personal identity; effectiveness and fitness of purpose; transparency of decisions; accountability for decisions; awareness of and guarding against misuse; and competence for 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 organisation, IEEE's report is one of the few that highlights effectiveness and competence as principles.

Download here

5. Montréal Declaration for Responsible Development of Artificial Intelligence (University of Montreal)

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 multistakeholder 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 operationalise 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.

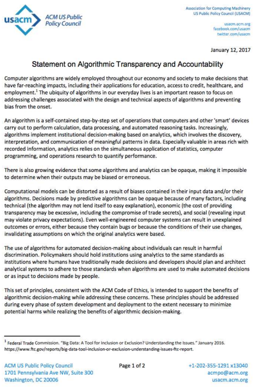
Download here

6. Statement on Algorithmic Transparency and Accountability (Association for Computing Machinery)

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

provenance; auditability; validation and testing

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; explaination of algorithmic decision-making; clarity of data provenance; auditability; and validation and testing, particularly against discrimination.



usacm ACM US Public Policy Council

Association for Computing Machinery
US Public Policy Council (USACM)

usacm.acm.org
facebook.com/usacm
twitter.com/usacm

January 12, 2017

Statement on Algorithmic Transparency and Accountability

Computer algorithms are widely employed throughout our economy and society to make decisions that have far-reaching impacts, including their applications for education, access to credit, healthcare, and employment.[1] The ubiquity of algorithms in our everyday lives is an important reason to focus on addressing challenges associated with the design and technical aspects of algorithms and preventing bias from the onset.

An algorithm is a self-contained step-by-step set of operations that computers and other 'smart' devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.

There is also growing evidence that some algorithms and analytics can be opaque, making it impossible to determine when their outputs may be biased or erroneous.

Computational models can be distorted as a result of biases contained in their input data and/or their algorithms. Decisions made by predictive algorithms can be opaque because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic (the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations). Even well-engineered computer systems can result in unexplained outcomes or errors, either because they contain bugs or because the conditions of their use changes, invalidating assumptions on which the original analytics were based.

The use of algorithms for automated decision-making about individuals can result in harmful discrimination. Policymakers should hold institutions using analytics to the same standards as institutions where humans have traditionally made decisions and developers should plan and architect analytical systems to adhere to those standards when algorithms are used to make automated decisions or as input to decisions made by people.

This set of principles, consistent with the ACM Code of Ethics, is intended to support the benefits of algorithmic decision-making while addressing these concerns. These principles should be addressed during every phase of system development and deployment to the extent necessary to minimize potential harms while realizing the benefits of algorithmic decision-making.

[1] Federal Trade Commission. “Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues.” January 2016. https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report.

ACM US Public Policy Council
1701 Pennsylvania Ave NW, Suite 300
Washington, DC 20006

Page 1 of 2

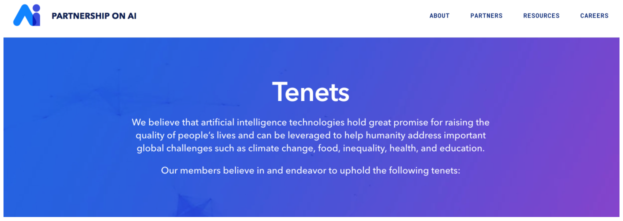
+1-202-355-1291 x13040
acmpo@acm.org
usacm.acm.org

Download here

7. Tenets (Partnership on AI)

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.



Tenets

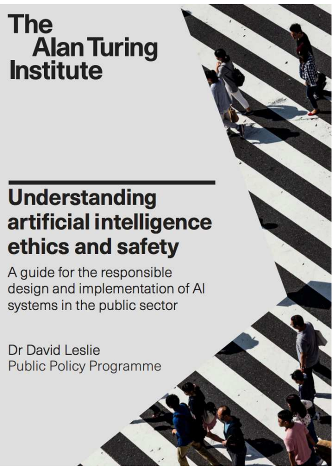
We believe that artificial intelligence technologies hold great promise for raising the quality of people’s lives and can be leveraged to help humanity address important global challenges such as climate change, food, inequality, health, and education.

Our members believe in and endeavor to uphold the following tenets:

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 maximise 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

8. Ethical Platform for the Responsible Delivery of an AI Project (The Alan Turing Institute)

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



The Alan Turing Institute

Understanding artificial intelligence ethics and safety

A guide for the responsible design and implementation of AI systems in the public sector

Dr David Leslie
Public Policy Programme

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.

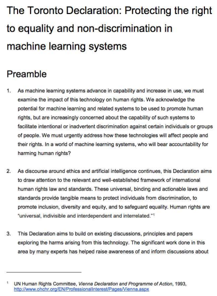
The report envisions a three-layered 'ethical platform', or a governance architecture, for an AI-power project in public service. The goal of this 'ethical platform' would essentially be to ensure that the project will uphold the following principles: ethically permissible in terms of the impacts it 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 behaviours.

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

Download here

9. The Toronto Declaration (Amnesty International and Access Now)

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



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

Preamble

1. As machine learning systems advance in capability and increase in use, we must examine the impact of this technology on human rights. We acknowledge the potential for machine learning and related systems to be used to promote human rights, but are increasingly concerned about the capability of such systems to facilitate intentional or inadvertent discrimination against certain individuals or groups of people. We must urgently address how these technologies will affect people and their rights. In a world of machine learning systems, who will bear accountability for harming human rights?

2. As discourse around ethics and artificial intelligence continues, this Declaration aims to draw attention to the relevant and well-established framework of international human rights law and standards. These universal, binding and actionable laws and standards provide tangible means to protect individuals from discrimination, to promote inclusion, diversity and equity, and to safeguard equality. Human rights are "universal, indivisible and interdependent and interrelated."[1]

3. This Declaration aims to build on existing discussions, principles and papers exploring the harms arising from this technology. The significant work done in this area by many experts has helped raise awareness of and inform discussions about

[1] UN Human Rights Committee, *Vienna Declaration and Programme of Action*, 1993, http://www.ohchr.org/EN/ProfessionalInterest/Pages/Vienna.aspx

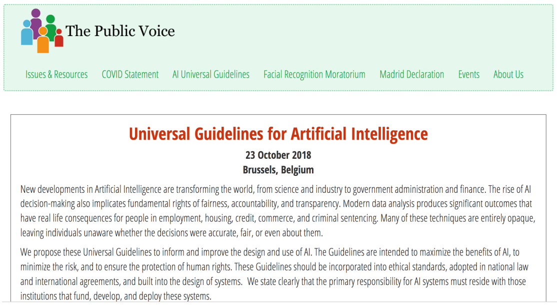
The Declaration is framed as a way to affirm the existing obligations and responsibilities of both states and private sector actors to promote, protect and respect human rights, as applied to the field of AI. As a corollary, in line with human rights law, it suggests also that use of systems must be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

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 organisations 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

10. Universal Guidelines for AI (The Public Voice Commission)

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



Universal Guidelines for Artificial Intelligence

23 October 2018
Brussels, Belgium

New developments in Artificial Intelligence are transforming the world, from science and industry to government administration and finance. The rise of AI decision-making also implicates fundamental rights of fairness, accountability, and transparency. Modern data analysis produces significant outcomes that have real life consequences for people in employment, housing, credit, commerce, and criminal sentencing. Many of these techniques are entirely opaque, leaving individuals unaware whether the decisions were accurate, fair, or even about them.

We propose these Universal Guidelines to inform and improve the design and use of AI. The Guidelines are intended to maximize the benefits of AI, to minimize the risk, and to ensure the protection of human rights. These Guidelines should be incorporated into ethical standards, adopted in national law and international agreements, and built into the design of systems. We state clearly that the primary responsibility for AI systems must reside with those institutions that fund, develop, and deploy these systems.

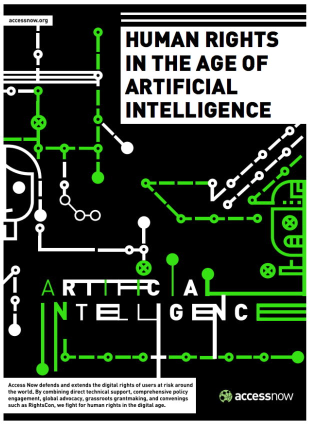
The Guidelines are meant to maximise the benefits of AI, minimise its risks, and ensure the protection of human rights. The document explicitly states the responsibility for these guidelines lies with the institutions that fund, develop and deploy these systems, and is fairly specific in describing what 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

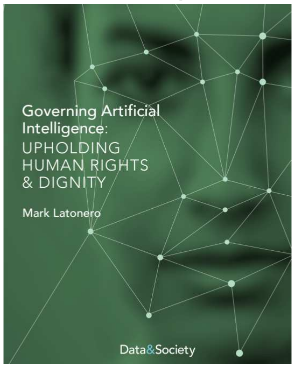
11. Human Rights in the Age of AI (Access Now)

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.


accessnow.org
HUMAN RIGHTS IN THE AGE OF ARTIFICIAL INTELLIGENCE
ARTIFICIAL INTELLIGENCE
Access Now defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age.
accessnow

The report describes some human rights which are impacted by AI, nothing 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 accountablility should concretely guide government and private sector use of AI.
Download here

12. Governing Artificial Intelligence. Upholding Human Rights & Dignity (Data & Society)


Governing Artificial Intelligence:
UPHOLDING HUMAN RIGHTS & DIGNITY
Mark Latonero
Data&Society

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.

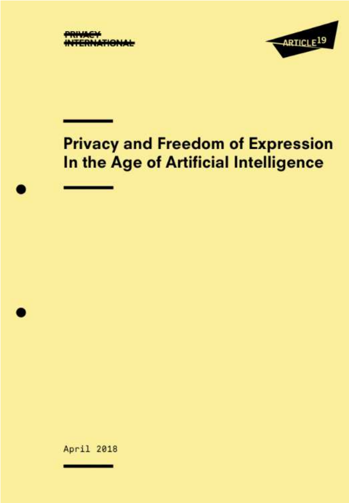
The report strongly recommends that the effects of AI on human rights should be constantly monitoring, and that human rights should not be seen as an ethical preference but as fundamental rights that should be enforced through law and regulation and supported by market incentives, public awareness and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

13. Privacy and Freedom of Expression in the Age of Artificial Intelligence (Privacy International & Article 19)

Key words	Developed by	Year
Human Rights; transparency; explainability; accountability	Access Now	Nov 2018

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 organisations 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.



PRIVACY INTERNATIONAL

ARTICLE 19

Privacy and Freedom of Expression
In the Age of Artificial Intelligence

April 2018

Download here

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

Key words	Developed by	Year
Gender equality; meaningful inclusion; digital equality, non-discrimination, open and transparent	W20	Oct 2018

W20 ARGENTINA 2018 WOMEN 20

POLICY BRIEF W20 ARGENTINA

Artificial Intelligence: open questions about gender inclusion

DIGITAL INCLUSION CO-CHAIR

WORLD WIDE WEB FOUNDATION

This policy brief provides concrete recommendations to mitigate the challenges of AI related to gender (design, deployment, and collateral effects of digitalisation 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 so as to allow for monitoring. Lastly, it recommends that research should be conducted to assessment the effects of AI on women's lives.

Download here

15. **Big Data and AI Principles in Engineering (WFEO)**

Developed by WFEO, **Year** Mar 2020 **Key Words** 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

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

16. AI Now 2019 Report

Developed by The AI Now Institute at New York University, **Year**: 2019

2019 Report

AINOW December 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]

17. White Paper from Wonks and Techies at Stanford University

Developed by a multidisciplinary group at Stanford University, cooperating on international technology and policy issues, led by Ms. Marietje Schaake, **Year**: June 2020

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 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. We submit these recommendations for your consideration, and look forward to the European Commission's comments."

[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 Organizations (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 theme for ITU's AI for Good Global Summit in May 2019 was "Accelerating Progress toward the SDGs."

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. The U.S. National Institute of Standards and Technology (NIST) AI standards strategy discussed in the next section includes an annex (Annex II) with a list of AI 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 Institut 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.

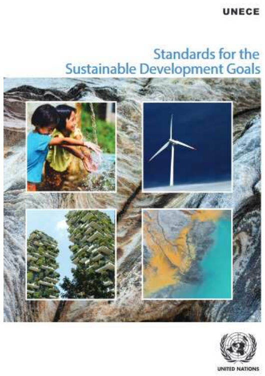
The NIST plan calls for standardization projects that lead to "... globally relevant and non-discriminatory standards, where standards avoid becoming non-tariff trade barriers or locking in particular technologies or products." Also so as not to stifle innovation, standards should have maximum flexibility, 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.[12]

Key References

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

1 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

[12] Citations for this chapter:

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

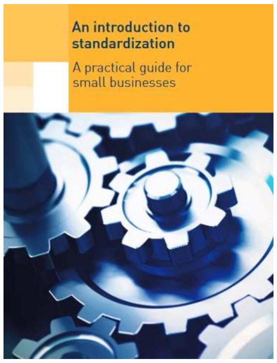
Artificial intelligence. (2020, September 02). National Institute of Standards and Technology. Retrieved from https://www.nist.gov/artificial-intelligence also be used in the regulations of products. For example, products that support the achievement of SDG 7 by improving energy efficiency.

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

2 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.


An introduction to standardization
A practical guide for small businesses

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

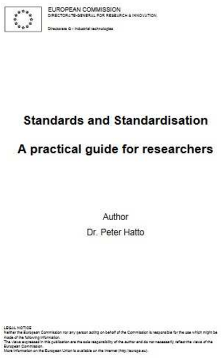
How standards are developed is the subject of the second section. Standards developed by the German Institute for Standardization (DIN) are designated DIN. DIN standards can result from work at the national, regional, or international level and are designated accordingly. Examples are given for the different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

3. 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 as a way for disseminating research results.

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 also 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.


EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION

Directorate G - Industrial technologies

Standards and Standardisation

A practical guide for researchers

Author
Dr. Peter Hatto

LEGAL NOTICE
Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.
The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.
More information on the European Union is available on the Internet (http://europa.eu).

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](#)

4. 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".[13]

NIST Special Publication 2000-01

ABC's of Conformity Assessment

Lisa Carnahan
Amy Phelps

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.2000-01

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

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,

[13] 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

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

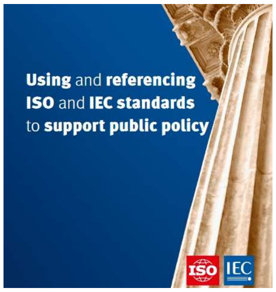
Conformity assessment procedures are important for global trade. The World Trade Organization (WTO) Technical Barriers to Trade (TBT) Agreement contains obligations regarding conformity assessment procedures and their use in international trade.

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 "cooperations" and established "multilateral agreements or arrangements" to recognize and accept the results of conformity assessment.

Download here

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

5 Using and referencing ISO and IEC standards to support public policy



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. A second is because of the alignment of good policy making and 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.[15] 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.[16] Section seven is for standards supporting public

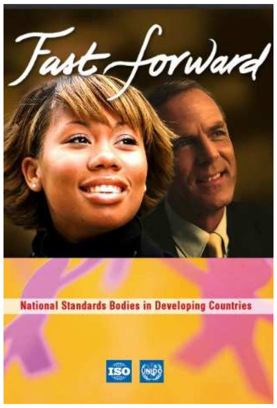
[15] ISO IEC Annex SO of the Consolidated ISO Supplement to the ISO/IEC Directives Part 1 https://www.iso.org/sites/directives/current/consolidated/index.xhtml#_idTextAnchor618 Accessed May 19, 2020

[16] 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.

policy in different sectors. A list is given of the sectors but readers are referred to a website for the examples.[17] [18]

Download here

6 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).

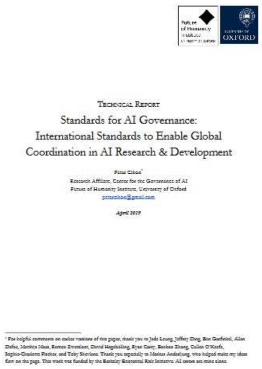
Part one outlines in some detail, the three pillars of a quality infrastructure system: metrology, standardization, and conformity assessment and accreditation. It is noted that having a fully functional quality infrastructure system may be beyond the resources of some developing countries. Some different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

[17] ISO Examples by Sector https://www.iso.org/sites/policy/sectorial_examples.html Accessed May 19, 2020
[18] IEC Examples by Product Sector https://www.iec.ch/perspectives/government/sectors/ Accessed May 19, 2020

7 TECHNICAL REPORT Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development



TECHNICAL REPORT

Standards for AI Governance:
International Standards to Enable Global
Coordination in AI Research & Development

Peter Cihon*
Research Affiliate, Center for the Governance of AI
Future of Humanity Institute, University of Oxford
petercihon@gmail.com

April 2019

* For helpful comments on earlier versions of this paper, thank you to Jade Leung, Jeffrey Ding, Ben Garfinkel, Allan Dafoe, Matthijs Maas, Remco Zwetsloot, David Hagebölling, Ryan Carey, Baobao Zhang, Cullen O'Keefe, Sophie-Charlotte Fischer, and Toby Shevlane. Thank you especially to Markus Anderljung, who helped make my ideas flow on the page. This work was funded by the Berkeley Existential Risk Initiative. All errors are mine alone.

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 remains a risk standards could fail to address important policy objectives and that AI research organizations that could contribute may not participate in standardization efforts.

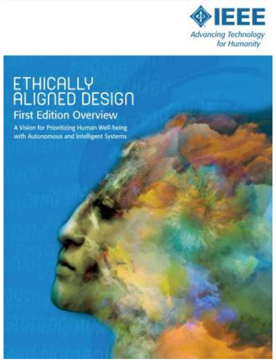
Standards cannot support all AI policy goals, but they are an important part of effective global solutions. Standards influence the development and deployment of AI systems through product specifications for example, explainability, robustness, and fail-safe design. They can also affect the larger context 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


IEEE
Advancing Technology for Humanity
ETHICALLY ALIGNED DESIGN
First Edition Overview
A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems

8 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^{TM} series of standardization projects explicitly focuses 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.

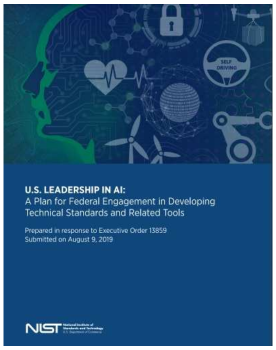
An A/IS Ethics Glossary has been developed and is available for download on the IEEE website. The glossary is intended for a broad audience of stakeholders including engineers, policymakers, philosophers, AI researchers, and standards developers.

As mentioned earlier, the IEEE P7000^{TM} 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.[19]
Download here

[19] IEEE Ethics in Action https://ethicsinaction.ieee.org/

9 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.



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

Prepared in response to Executive Order 13859
Submitted on August 9, 2019

SELF DRIVING
NIST National Institute of Standards and Technology U.S. Department of Commerce

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.

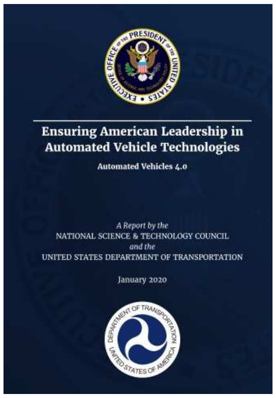
Specific recommendations for the Federal government are offered on coordination, research, partnerships and international engagement.

Download here

10 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.



Ensuring American Leadership in Automated Vehicle Technologies

Automated Vehicles 4.0

A Report by the NATIONAL SCIENCE & TECHNOLOGY COUNCIL and the UNITED STATES DEPARTMENT OF TRANSPORTATION

January 2020

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 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

11 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.

FINAL REPORT

An Artificial Intelligence Standards Roadmap: Making Australia's Voice Heard

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

Chapter 4: National AI Strategies and International Organizations

Introduction

Followed by a brief discussion in previous chapters, how country reacts to the development of AI and how country supports the development of AI is going to have a far-reaching impact on its own competitiveness worldwide. Therefore, countries start to announce national strategies on AI successively to contest for the leadership on AI development. This chapter is designed to provide an overview of the global landscape of different national AI strategies and the feature of these documents. It is notable that those AI strategies put in place in not only to strengthen the development of AI, but also drive the development of other related industrial, then to the overall economy, as well as to standardize the application of AI for regulations, frameworks to address ethical concerns. This chapter presents a compilation of national strategies on AI and those from other stakeholders.

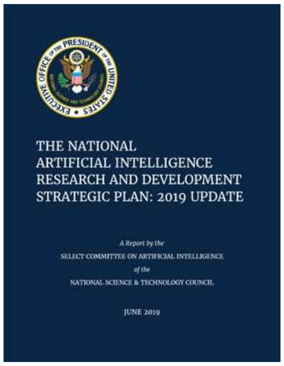
Key References

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

National Strategies

1. 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.



THE NATIONAL
ARTIFICIAL INTELLIGENCE
RESEARCH AND DEVELOPMENT
STRATEGIC PLAN: 2019 UPDATE

A Report by the
SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE
of the
NATIONAL SCIENCE & TECHNOLOGY COUNCIL

JUNE 2019

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

2. 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

The Trump Administration is committed to promoting an international environment that supports AI R&D and opens markets for American AI industries while also ensuring that the technology is developed in a manner consistent with our Nation's values and interests.

Download here

3. 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.

Die Bundesregierung

KI Nationale Strategie für Künstliche Intelligenz AI Made in Germany

Strategie Künstliche Intelligenz der Bundesregierung

Stand: November 2018

www.ki-strategie-deutschland.de

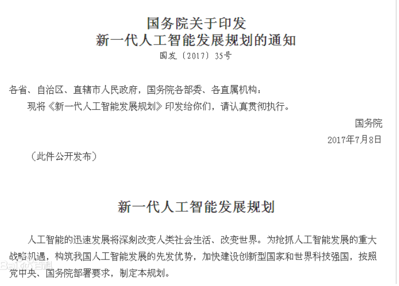
The Federation allocated a total of €500 million to beef up the AI strategy for 2019 and the following years. Up to and including 2025, the Federation intends to provide around €3 billion for the implementation of the Strategy.

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

4. 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:



国务院关于印发
新一代人工智能发展规划的通知
国发〔2017〕35号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

现将《新一代人工智能发展规划》印发给你们，请认真贯彻执行。

国务院
2017年7月8日

（此件公开发布）

新一代人工智能发展规划

人工智能的迅速发展将深刻改变人类社会生活、改变世界。为抢抓人工智能发展的重大战略机遇，构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国，按照党中央、国务院部署要求，制定本规划。

(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.

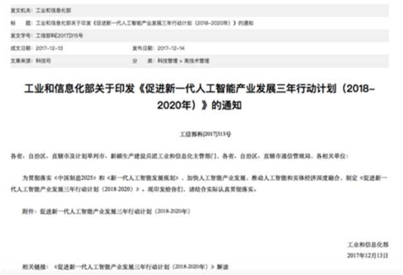
(3) by 2030, China's AI theories, technologies, and applications should achieve world-leading levels, making China the world's primary AI innovation center, laying an important foundation for becoming a leading innovation-style nation and an economic power.

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

5. China: Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018–2020)



发文机关：工业和信息化部	
标　题：工业和信息化部关于印发《促进新一代人工智能产业发展三年行动计划（2018-2020年）》的通知	
发文字号：工信部科〔2017〕315号	
成文日期：2017-12-13	发布日期：2017-12-14
文章来源：科技司	分　类：科技管理 > 高技术管理

工业和信息化部关于印发《促进新一代人工智能产业发展三年行动计划（2018–2020年）》的通知

工信部科〔2017〕315号

各省、自治区、直辖市及计划单列市、新疆生产建设兵团工业和信息化主管部门，各省、自治区、直辖市通信管理局，各相关单位：

为贯彻落实《中国制造2025》和《新一代人工智能发展规划》，加快人工智能产业发展，推动人工智能和实体经济深度融合，制定《促进新一代人工智能产业发展三年行动计划（2018-2020）》，现印发给你们，请结合实际认真贯彻落实。

附件：促进新一代人工智能产业发展三年行动计划（2018-2020年）

工业和信息化部
2017年12月13日

相关链接：《促进新一代人工智能产业发展三年行动计划（2018-2020年）》解读

In order to implement the plans of *Made in China 2025* and *New Generation of Artificial Intelligence Development Plan*, the Ministry of Industry and Information Technology (MIIT) issued this *Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020)* in July 2017. It focuses on the in-depth integration of information technology and manufacturing technology to speed up the building of China into a manufacturing superpower and a cyber superpower. Through the implementation of four key tasks, China strives to achieve a major breakthrough in a series of landmark AI products by 2020, establish an international competitive advantage in several key areas.

(1) Scale-up the development of key AI products, including intelligent networked vehicles, intelligent service robots, intelligent unmanned aerial vehicles, medical imaging diagnosis systems, video image identification systems, intelligent voice interactive systems, intelligent translation systems and smart phone products.

(2) Significantly enhance core competencies in AI, including smart sensors, neural network chips and open-source platforms.

(3) Deepen the development of smart manufacturing, accelerate integrated applications of complex environment identification, new-type human-machine interaction, etc., AI technologies in key technical equipment. Improve the level of application of new models such as intelligent production, large-scale personalized customization, and predictive maintenance.

(4) Establish the foundation for an AI industry support system, including industry training resources, standard testing and intellectual property service platforms, intelligent network infrastructure and cybersecurity systems.

Download here

6. 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 documents sets out a European initiative on AI, with three strategic goals:

EUROPEAN
COMMISSION

Brussels, 25.4.2018
COM(2018) 237 final

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

{SWD(2018) 137 final}

(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

7. European Union: Coordinated Plan on Artificial Intelligence

EUROPEAN
COMMISSION

Brussels, 7.12.2018
COM(2018) 795 final

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

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; (8) International cooperation.

Download here

8. European Union: The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)

The Ethics Guidelines for Trustworthy Artificial Intelligence (AI) is a document prepared by the High-Level Expert Group on Artificial Intelligence (AI HLEG). This independent expert group was set up by the European Commission in June 2018, as part of the AI strategy announced earlier that year. The document provides 3 ethical principles, 7 requirements and a concrete and non-exhaustive assessment list towards Trustworthy AI.

INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION

AI

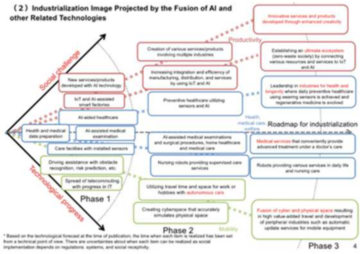
ETHICS GUIDELINES
FOR TRUSTWORTHY AI

The documents points out that develop, deploy and use AI systems should be in a way that adheres to the ethical principles of: respect for human autonomy, prevention of harm, fairness and explicability. Acknowledge and address the potential tensions between these principles. And the development, deployment and use of AI systems should meet the seven key requirements for Trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability. It's also important to adopt a Trustworthy AI assessment list when developing, deploying or using AI systems, and adapt it to the specific use case in which the system is being applied. But such an assessment list will never be exhaustive. Ensuring Trustworthy AI is not about ticking boxes, but about continuously identifying and implementing requirements, evaluating solutions, ensuring improved outcomes throughout the AI system's lifecycle, and involving stakeholders in this.

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 Abes for several years with a package of policies ("Abenomics").


(2) Industrialization Image Projected by the Fusion of AI and other Related Technologies

Social challenge
Technological progress
Productivity
Health, medical care welfare
Mobility
Roadmap for industrialization
Phase 1
Phase 2
Phase 3
4
New services/products developed with AI technology
IoT and AI-assisted smart factories
AI-aided healthcare
Health and medical data preparation
AI-assisted medical examination
Care facilities with installed sensors
Driving assistance with obstacle recognition, risk prediction, etc.
Spread of telecommuting with progress in IT
Creation of various services/products involving multiple industries
Increasing integration and efficiency of manufacturing, distribution, and services by using IoT and AI
Preventive healthcare utilizing sensors and AI
AI-assisted medical examinations and surgical procedures, home healthcare and medical care
Nursing robots providing supervised care services
Utilizing travel time and space for work or hobbies with autonomous cars
Creating cyberspace that accurately simulates physical space
Innovative services and products developed through enhanced creativity
Establishing an ultimate ecosystem (zero-waste society) by connecting various resources and services to IoT and AI
Leadership in industries for health and longevity where daily preventive healthcare using wearing sensors is achieved and regenerative medicine is evolved
Medical services that conveniently provide advanced treatment under a doctor's care
Robots providing various services in daily life and nursing care
Fusion of cyber and physical space resulting in high value-added travel and development of peripheral industries such as automatic update services for mobile equipment

* Based on the technological forecast at the time of publication, the time when each item is realized has been set from a technical point of view. There are uncertainties about when each item can be realized as social implementation depends on regulations, systems, and social receptivity.

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

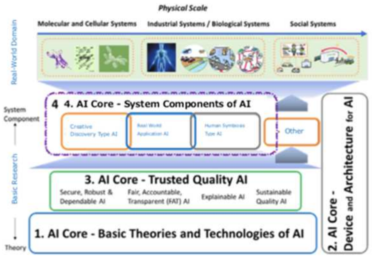

Physical Scale
Molecular and Cellular Systems
Industrial Systems / Biological Systems
Social Systems
Real-World Domain
System Component
Basic Research
Theory
4 4. AI Core - System Components of AI
Creative Discovery Type AI
Real World Application AI
Human Symbiosis Type AI
Other
3. AI Core - Trusted Quality AI
Secure, Robust & Dependable AI
Fair, Accountable, Transparent (FAT) AI
Explainable AI
Sustainable Quality AI
1. AI Core - Basic Theories and Technologies of AI
2. AI Core - Device and Architecture for AI

Figure: Overall Structure of AI R&D

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.

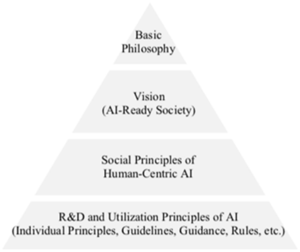
To support these strategic objectives, this Strategy establishes an integrated policy package for AI that encompasses educational reform, research and development (R&D), social implementation, data-related infrastructure construction and AI era digital government in order to contribute to the world, overcome challenges, and ultimately improve Japan's industrial competitiveness.

Download here

11. Japan: Social Principles of Human-Centric AI

In March of 2019, Japan government compiled the "Social Principles of Human-Centric AI."

This document specifies the form of society that Japan should aim for, a multilateral framework, and a policy direction that the national and local governments should aim for as AI develops.

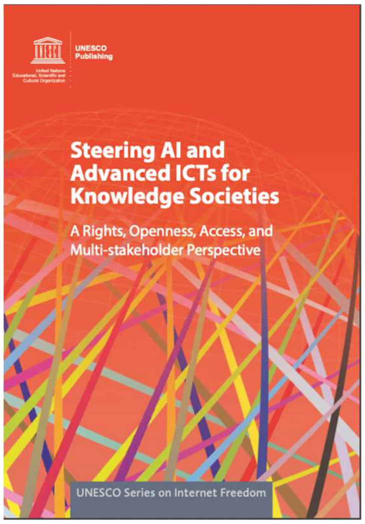

Basic Philosophy
Vision (AI-Ready Society)
Social Principles of Human-Centric AI
R&D and Utilization Principles of AI (Individual Principles, Guidelines, Guidance, Rules, etc.)

It defines three points as its basic philosophy: (1) Dignity - A society in which human dignity is respected; (2) Diversity and Inclusion - A society in which people with diverse backgrounds can pursue their own well-being; (3) Sustainability - A sustainable society.

It sets out 7 social principles of AI: (1) the human-centric principle; (2) the principle of education/literacy; (3) the principle of privacy protection; (4) the principle of ensuring security; (5) the principle of fair competition; (6) the principle of fairness, accountability, and transparency; (7) the principle of innovation.

Download here

Steering AI and Advanced ICTs for Knowledge Societies: A ROAM Perspective, 2019


UNESCO Publishing
United Nations Educational, Scientific and Cultural Organization
Steering AI and Advanced ICTs for Knowledge Societies
A Rights, Openness, Access, and Multi-stakeholder Perspective
UNESCO Series on Internet Freedom

There are no simple answers about what the future holds for humanity, this 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-centred 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

Chapter 5: Conclusion and the Way Forward

Artificial Intelligence (AI) has become a popular topic in recent years; it has captured great attention not only from technology sectors but also from governments, academia, and the general public, as AI poses broad impacts on society, economy, and environment. Artificial Intelligence is not a new technology; it was first introduced in the early 1950s, notably in the Dartmouth Summer Research Project on Artificial Intelligence in 1956. AI has been through more than 60 years of development, and with the help of recent advancements in technology, economy, and IT infrastructure, AI has broken through historical technical limitations. AI has now moved to a new era and put in place for better applications on daily production. This section aims to provide a general understanding of AI, including its applications, impacts on society, economy and environment, and positioning.

Artificial Intelligence can be simply understood 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, learning, etc. AI technology has been rapidly developing, with some AI having 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 chess, and even beat the world Go champion, Lee Sedol. Artificial Intelligence has been integrated into our daily lives for a long time. Smartphones, Google searches, face recognition, call center voice assistance and many other services that we already use are actually examples of AI technologies. For higher-level applications, AI can be applied in healthcare diagnoses, financial risk management, agricultural monitoring, autonomous driving, and many other purposes to be developed and discovered. AI is a broad, general, and complicated technology that requires multi-dimensional knowledge, research, and development.

The fast development of AI contributes numerous positive benefits to our society, economy and environments, and it certainly advances the achievement of Sustainable Development Goals. AI creates significant possibilities for the business production process. In manufacturing, for instance, AI drives the application of robotics, which reduces unnecessary production costs for wasted material, time, and labor. According to available data, this in turn leads to greater productivity and economic growth.

Frontier technologies such as AI, Big Data, Internet of Things and many others no longer exist as independent technologies. Every technology plays a role in supporting the production of others, giving prominence to the collaboration between technologies. The

Internet of Things (IoT) allows people to monitor property, production, and other ongoing activities remotely, which can be utilized in marine management, agriculture, households, and more. Combined use of IoT and AI can utilize data for effective control of sustainable production to reduce CO2 emission. AI improves the effectiveness of public services and enhances government accountability and transparency. AI makes education and healthcare more accessible, which helps reduce inequality in the long term. Many further benefits are yet to be discovered.

However, the proliferation of AI will present challenges, such as unemployment, which will be more notable in developing countries and potentially escalate economic and gender inequalities. Ethical concerns such as data protection, privacy, ownership of intellectual property, legal rights, rules and regulation for autonomous driving, along with others are urgently needed to be solved. We must take measures to prevent these technological advances from possibly exacerbating challenges or creating new ones. The emergence of AI does not necessarily signify the replacement of low-skill jobs. On the contrary, a new form of the industrial model can be established, enabling the creation of many new jobs. Personal data and privacy can be more protected with corresponding regulation put in place. This requires effective policymaking from the local government, high-quality education for everyone, as well as regular strong and continuous collaborations between organizations and industries.

The development and use of AI technologies will continue to engender transformations in society, which 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 requires continuous research, multi-stakeholder cooperation and periodic updating of the ethical standards that guide the development and use of AI. The UN system is uniquely positioned to provide a platform to facilitate standard setting, exchange of knowledge, and cooperation among different stakeholder groups from the Global North and South.

The resources provided in this Guide provide an overview of global discussions on the Ethics of AI. These discussions are laying the roadmap for cooperation in ethically-informed governance of AI by articulating international, regional, and national agenda opportunities and concerns related to 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 are examples of processes to facilitate cooperation on the Ethics of AI within the UN system and its Member States.

The resources highlighted in this Guide showcased that standards and conformity assessment already play an essential role in supporting the achievement of the SDGs. Policymakers need to understand the important role that standards and conformity assessment play and could play in governance and in enabling the digitalization that is necessary to achieve the SDGs. Standardization and conformity assessment are tools that need to be and can be used effectively at all governance levels from local to global. Education and capacity building on standardization and conformity assessment should be encouraged and supported utilizing the many resources already available.

The standards community and the technical community in general need to be made more aware of the SDGs and how technology, standardization, and conformity assessment, specifically can contribute to the SDGs. There needs to be a much broader effort to engage the technical community on the SDGs. There also needs to be broader involvement in standardization activities by the technical community, including AI academic and industrial researchers.

Smartphones are an integral part of today's technological landscape and depend on an extensive base of technical standards from the device level to the global internet. The existing standards and conformity assessment system supports this and AI standardization and conformity assessment should build on this. One of the challenges of standardization generally is when and if to standardize so as not to stifle innovation or lock in the wrong technology. AI standardization is in its very early stages and there is broad support for international standardization that first involves building understanding with all technical and non-technical stakeholders.

In our increasingly digital and data-driven world, AI may very well bring on a new era. Undoubtedly, AI may be a force for good, enabling groundbreaking insights and applications. Yet AI evidently poses significant risks as well. As we face the exciting prospects that AI can usher in, especially in accelerating the achievement of the SDGs, 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, we must look at how the field has evolved, and where it stands today. The following **key observations** have been identified for policy makers:

Limited accessibility of AI research:

As the AI landscape is changing, there are several notable trends that have emerged. The 2020 State of AI Report finds that one of the major trends is the limited accessibility of AI research, in which little improvement has been made since 2016. Only 15% of AI research papers publish their code, many of which come from academic groups rather than industry. As such, there is very little accountability and a limited degree of reproducibility in AI, limiting progress. Organizations that do not publish their code include OpenAI and DeepMind, two of the largest tech companies.

However, there are rare but important exceptions, such as PostEra's COVID Moonshot initiative. Moonshot is an initiative in which an international team of scientists are working without intellectual property constraints, fully open source on a crowdsourced enterprise to accelerate the development of a COVID-19 antiviral. They use machine learning to determine which drug designs to make and test, accomplishing tasks in less than 48 hours that would take human chemists about a month. 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 analyse, understand and manipulate language, has been the focus of AI use today, predominantly taking place in large companies with huge models and astronomical training costs. While AI has traditionally had an open ethos, the industrialization of AI has been diminishing that (to retain their IP) and centralizing AI talent.

Corporate-driven brain drain*:*

This embodies another major trend reported in the 2020 State of AI Report: significant "corporate-driven brain drain." Companies including DeepMind, Amazon, and Microsoft 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.

Currently, top-tier AI research has been dominated by the United States, whose universities and corporations have been leading in acceptances of major academic conference papers. This lead is driven by international talent. The majority of principal AI researchers working in the US were trained elsewhere, commonly in China.

More private funding

Private funding for AI-first companies has also remained strong, with 2020 likely to see over $25 billion in total volume and 350+ deals. Public policies may further attract investment in AI. Several governments have taken steps to foster AI development, including offering tax breaks for AI entrepreneur immigrants and introducing special tech visas that do not require work sponsorship. Governments have also invested in funding for science programs and AI R&D and have the influence to boost startups as credible first customers. AI-friendly local regulations are also likely a motivator for scaling AI investment, as it ensures that new technologies can be piloted with protections of data ownership and clearly-defined legal liability. Settings where entrepreneurs feel safe to develop AI technologies draw AI investors.

Interestingly, the COVID-19 global pandemic has not had as detrimental an effect on AI investments as one may expect. According to a Gartner poll, 47% of investments in AI remained untouched, and 30% of organizational investors actually intended to increase their investments. Only 7% decreased AI investments and the remaining percent just temporarily suspended them.

As a result, there have been great strides in AI in both academia and in industry. Significant progress has begun to be realized in AI drug discovery. For example, Japan recently initiated its first phase in a clinical trial of an AI-designed drug to treat obsessive-compulsive disorder. One of the most popularly discussed applications of AI is in autonomous vehicles (AV). While there has been a surge in AV miles driven compared to prior years, observations from California show that the mileage from self-driving cars is still miniscule when compared to human-driven cars.

Growing ethical concerns:

National authorities and the technical community need to address the ethical aspects of designing and building AI systems and applications. Technical standards cannot address all the ethical implications of AI and digitalization but they will be an important component.

In just a few months of the COVID-19 crisis, hundreds of thousands have died, tens of millions pushed into poverty and hunger, and inequalities exacerbated along many dimensions. The pandemic has already demonstrated the promise and power of technology to alleviate its impacts. Simultaneously, it has heightened disquieting issues of lack of access and invasion of privacy.

The COVID-19 pandemic has also 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 build-up 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, artificial intelligence too, must be guided by human values, human solidarity and human intelligence.

There are substantial ethical ramifications and public concerns from the ever-increasing adoption of AI technologies. Among the various ethical risks that have arisen in the field of AI, issues from widespread facial recognition were given the spotlight in the 2020 State of AI Report. About half of the world currently permits the use of facial recognition, which is used in functions from unlocking a smartphone to statewide surveillance. Recently, Russia has been utilizing facial recognition technology to monitor whether quarantine mandates are being upheld by potential COVID-19 carriers. Only a few countries, such as Belgium, Luxembourg, and Morocco have partial bans on the technology.

Technical standards, such as specifying requirements on AI systems' 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 best 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.

In sum, there is a need to review both the policy and regulation framework 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, though many remain superficial. The requirements of transparency, auditability, 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.

With the climate crisis also upon us, we have no time to waste in charting the way forward. We need to harness the full power of game-changing technologies like AI if we are to make a decisive break towards a new, more sustainable and equitable, 'normal'.

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.

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 intends to review and propose a research agenda covering the main elements and key issues for policy development beyond the Guide's scope.

As AI continues to evolve and grow, it is essential that we understand the trends and where we are heading in order to best develop a safe, equitable, and beneficial path for AI implementation in the future. Further research and policy deliberation in this area would assist in filling knowledge gaps and contribute to achieving the 17 Sustainable Development Goals.[20]

[20] 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 Qarterly: 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 McKinsey & Company. Retrieved from https://www.weforum.org/whitepapers/the-next-economic-growth-engine-scaling-fourth-industrial-revolution-technologies-in-production.

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=E

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 Qarterly: 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_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.

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 CICG. 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 recommendation on the ethics of artificial intelligence

TIMEFRAME	ROADMAP ACTION
February-April, 2020	• Background research and document preparation for the first Ad Hoc Expert Group meeting
20-24 April 2020	• First meeting of the Ad Hoc Expert Group to prepare the draft text of the Recommendation
May-July 2020	• Open multi-stakeholder consultations at the national, regional and international levels on the draft text • Online consultations on the draft text
End August 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	• The Secretariat prepares a final report containing one or more draft texts of the Recommendation based on comments and observations from Member States
April 2021	• Transmission of the final report containing one or more draft texts of the Recommendation to Member States at least seven months before the General Conference • First session of the special committee of intergovernmental experts (category II meeting) to prepare a final draft of the Recommendation
June 2021	• Second session of the special committee of intergovernmental experts (category II meeting) 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 by the General Conference

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

ILO

- **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** 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.

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.
- **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.

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

- As a follow-up of World Summit on the Information Society's (WSIS), UNESCO has 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 has 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 new publication entitled **"Steering AI and Advanced ICTs for Knowledge Societies: a ROAM perspective"** was launched at the Internet Governance Forum in 2019.

- UNESCO's Information For All Programme (IFAP) examined and approved the **Code of Ethics for the Information Society**
- 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.**

  - **BACKGROUND ON THE UNESCO RECOMMENDATION**
    The UNESCO Recommendation on the Ethics of Artificial Intelligence will outline recommended principles and policy actions addressed primarily to Member States, as well as other stakeholders such as the private sector, civil society, etc. If the Recommendation is adopted, Member States will be invited to submit periodic reports (normally every four years) on the measures that they have adopted in relation to the Recommendation. This reporting modality also will serve as a monitoring mechanism to identify best practices, gaps, challenges for implementation, emerging risks and new principles that are needed as AI develops. Support will be provided to assist Member States on the implementation of the Recommendation as necessary, and as appropriate. In this regard, the Recommendation will be an opportunity for Member States to discuss and agree upon an initial non-exhaustive set of basic principles and recommended policy actions as ethical and human rights guardrails for the ethical design, development and deployment of AI.

- UNESCO has also organized a series of event 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)

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.

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 programme** 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 and,
    - UNICRI and INTERPOL will explore the development of a toolkit for the responsible use of AI by law enforcement in 2020

<u>UNSG's High Level Panel on Digital Cooperation</u>

- **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.

<u>UNU</u>

- 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 Al on IP policy. Outcome of the questionnaire may form the basis for future structured discussions.

Annex IV: Links to resources on ethics of AI

1. UNESCO:
   a. Preliminary Study on the Ethics of Artificial Intelligence (2019)
   b. Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective (2019)
   c. UNESCO's Internet Universality Indicators: A Framework for Assessing Internet Development (2019)
   d. Final Report on the International Conference on Artificial Intelligence and Education. Planning Education in the AI Era: Lead the Leap (2019)
   e. Beijing Consensus on Artificial Intelligence and Education (2019)
   f. I'd blush if I could: closing gender divides in digital skills through education (2019)
   g. Two-Eyed AI: A Reflection on Artificial Intelligence (2019)
   h. Bangkok Statement on the Ethics of Science and Technology and Sustainable Development (2019)
   i. Human Decisions: Thoughts on AI (2018)
   j. Report of COMEST on Robotics Ethics (2017)
2. The United Nations System:
   a. The Age of Digital Interdependence, Report of the UN Secretary-General's High-level Panel on Digital Cooperation (2019)
   b. The right to privacy in the digital age. Resolution adopted by the Human Rights Council (2019)
   c. Policy Inputs for the Young UN Policy Lab (2018)
   d. AI for Good Global Summit Report, ITU (2017)

3. 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)
      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)

4. 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. France: Strategy for a Meaningful Artificial Intelligence (2018)

d. Germany:

i. Opinion of the Data Ethics Commission (2019)

ii. Automated and Connected Driving, BMVI Ethics Commission report (2017)

e. 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)

f. New Zealand:

i. Government Use of Artificial Intelligence in New Zealand (2018)

g. 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)

h. Sweden:

i. Artificial Intelligence in Swedish Business and Society (2018)

i. United Kingdom:

i. Code of conduct for data-driven health and care technology, Department of Health and Social Care (2019)

5. 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)

6. 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)

Annex V: Summary Table of Ethical Principles

	Principle	*Key message*	*Sources*
Potentially relevant principles			
1	**Human rights**[21]	AI should be developed and implemented in accordance with international human rights standards.	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	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 in order to be effective and efficient tools or mediators.	COMEST 2017
	Rights-based	[Internet] rooted in the Universal Declaration of Human Rights and its associated Covenants.	UNESCO General Conference 2015 decision

[21] Suggested by this background document as a foundational value in a different formulation.

	Principle	***Key message***	***Sources***
			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, 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.	G20 AI Principles 2019 = OECD AI Principles 2019
	Human rights	Ensure autonomous and intelligent systems do not infringe on internationally recognised 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

	Principle	*Key message*	*Sources*
2	**Inclusiveness**[22]	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 sector, international organizations, civil society and academia.	CEB 2019

[22] Suggested by this background document as a foundational value in a different formulation.

	Principle	*Key message*	*Sources*
	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

	Principle	*Key message*	*Sources*
	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, sustainable development, peace and prosperity.	CEB 2019
	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
	Prioritising well-being	Prioritise metrics of well-being in the design and use of AISs 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	Berkman Klein Center 2020

	Principle	***Key message***	***Sources***
		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.	
	Beneficence	While promoting good is often mentioned, it is rarely defined, though notable exceptions mention the *augmentation of human senses, the promotion of human well-being and flourishing, peace and happiness*, the creation of socio-economic opportunities, and *economic prosperity*. 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, privacy, and cultural and gender diversity, as well as fundamental human rights.	UNI Global Union 2017
	Well-being	AIs 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.	COMEST 2019

	Principle	***Key message***	***Sources***
		NB: need to take into account situations where human control could be detrimental.	
	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 standardise 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

	Principle	*Key message*	*Sources*
	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 control over, and responsibility for, these machines at all times.	UNI Global Union 2017
	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.	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: 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; and, 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,	Berkman Klein Center

	Principle	*Key message*	*Sources*
		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.	
	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.	The global landscape of AI ethics guidelines, Nature 2019

	Principle	***Key message***	***Sources***
		References to transparency comprise efforts to *increase explainability, interpretability or other acts of communication and disclosure*. 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.	
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	Minimise 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 of responsibility	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

	Principle	*Key message*	*Sources*
	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** *(often cited in combination with responsibility or used interchangeably)*	Arrangements should be developed that will make possible to attribute accountability for AI-driven decisions and the behaviour 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	WEF Principles Development Tool 2020

	Principle	***Key message***	***Sources***
		ethical principles when working with AI and be held accountable when AIs break ethical principles and voluntary obligations.	
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 organisations and trade unions.	UNI Global Union 2017

	Principle	*Key message*	*Sources*
12	**Sustainability**[23]	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 so as 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 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.	The global landscape of AI ethics guidelines, Nature 2019
Other relevant principles or sub-principles			
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

[23] Suggested by this background document as a foundational value in a different formulation.

	Principle	*Key message*	*Sources*
	'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	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 outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.	G20 AI Principles 2019 = OECD AI Principles 2019

	Principle	*Key message*	*Sources*
		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.	
	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 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

	Principle	*Key message*	*Sources*
		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	COMEST 2017

	Principle	*Key message*	*Sources*
		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.	
	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 *data protection and data security*. A few sources link privacy to freedom or trust.	The global landscape of AI ethics guidelines, Nature 2019
	Privacy	AIs and people with AI responsibilities should protect personal and client data. Those who gather or share data with AIs or from AIs 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

	Principle	*Key message*	*Sources*
	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 favour of "radical individualism".	The global landscape of AI ethics guidelines, Nature 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 emphasise 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)	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. The value of justice is also related to non-discrimination. Roboticists should be sensitised to the reproduction of gender bias	COMEST 2017

	Principle	*Key message*	*Sources*
		and sexual stereotype in robots. The issue of discrimination and stigmatisation through data mining collected by robots is not a trivial issue. Adequate measures need to be taken by States.	
	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 ne*ed 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 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	AIs 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, decent work, technical and socioeconomic challenges.	CEB 2019

	Principle	*Key message*	*Sources*
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 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 experimentation99, manipulation or surveillance. Freedom and autonomy are believed to be promoted through transparency and predictable AI55, by not "reducing options for and knowledge of citizens", by actively increasing people's knowledge about AI, giving notice and consent	The global landscape of AI ethics guidelines, Nature 2019

	Principle	*Key message*	*Sources*
		or, conversely, by actively refraining from collecting and spreading data in absence of informed consent.	
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 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.	The global landscape of AI ethics guidelines, Nature 2019
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	**Professionalism**	AI researchers, scientists and technicians should follow high scientific and professional standards.	WEF Principles Development Tool 2020

Note: UNESCO.

Annex VI 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.

	Key Understanding	International Organization	Private Sectors	Academia
AI	The brain of machine	Refer computer systems can preform 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 as known as Artificial Narrow Intelligence(ANI) or Narrow AI. It refers to a AI is designed to perform one specific or small set of tasks such as facial ID recognition,	Design to perform narrow task, e.g., play chess, facial recognition, internet search, driving car. (ESCAP,2018)		

	voice-activated assistants, autonomous driving etc. It is notable that weak AI has possessed the ability to surpass human such chess playing.			
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.			
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		It is based on algorithms that can learn from data	

	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.		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 is 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 posses similar intelligence as human.			