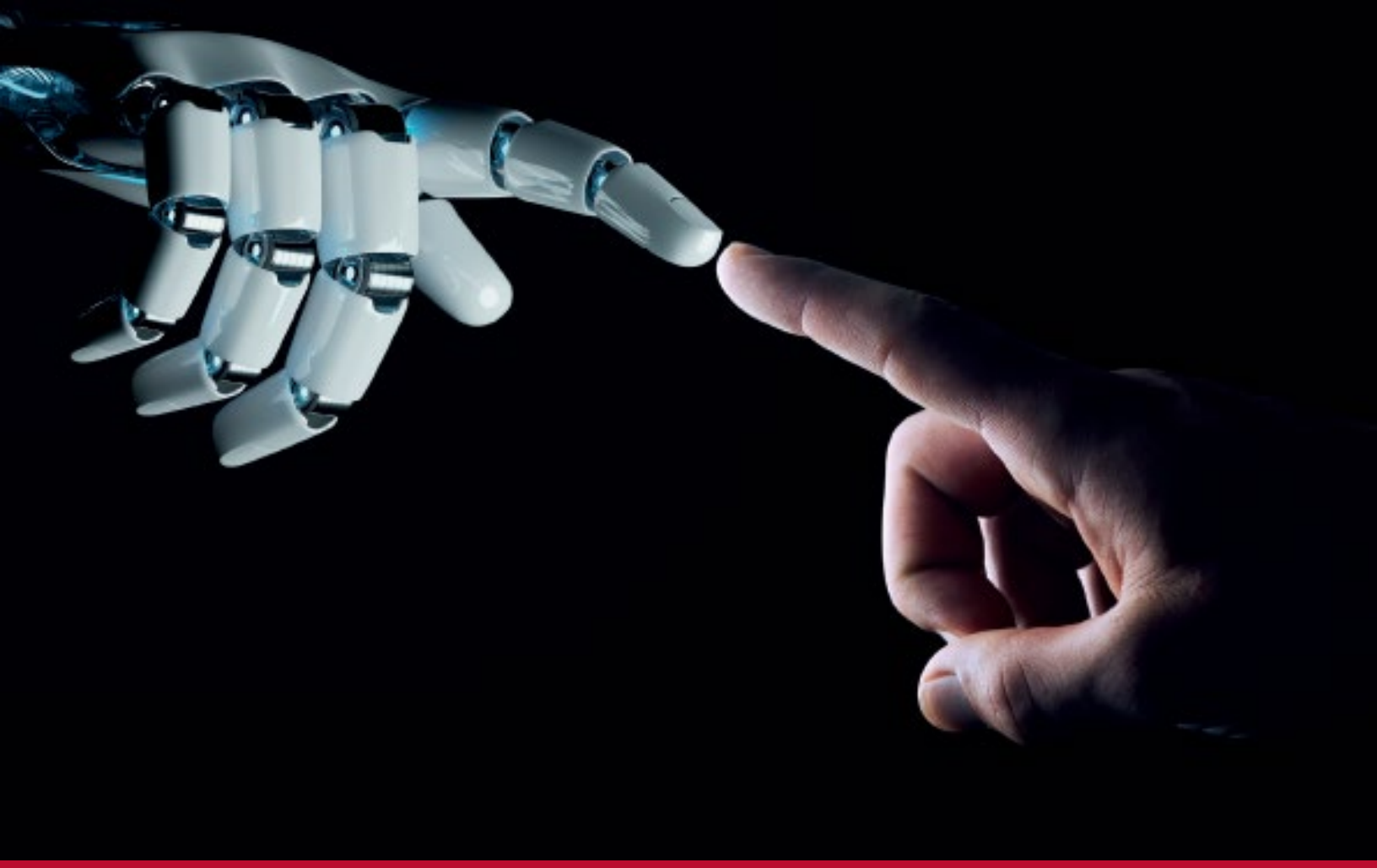
USAID
FROM THE AMERICAN PEOPLE



ARTIFICIAL INTELLIGENCE (AI) ETHICS GUIDE

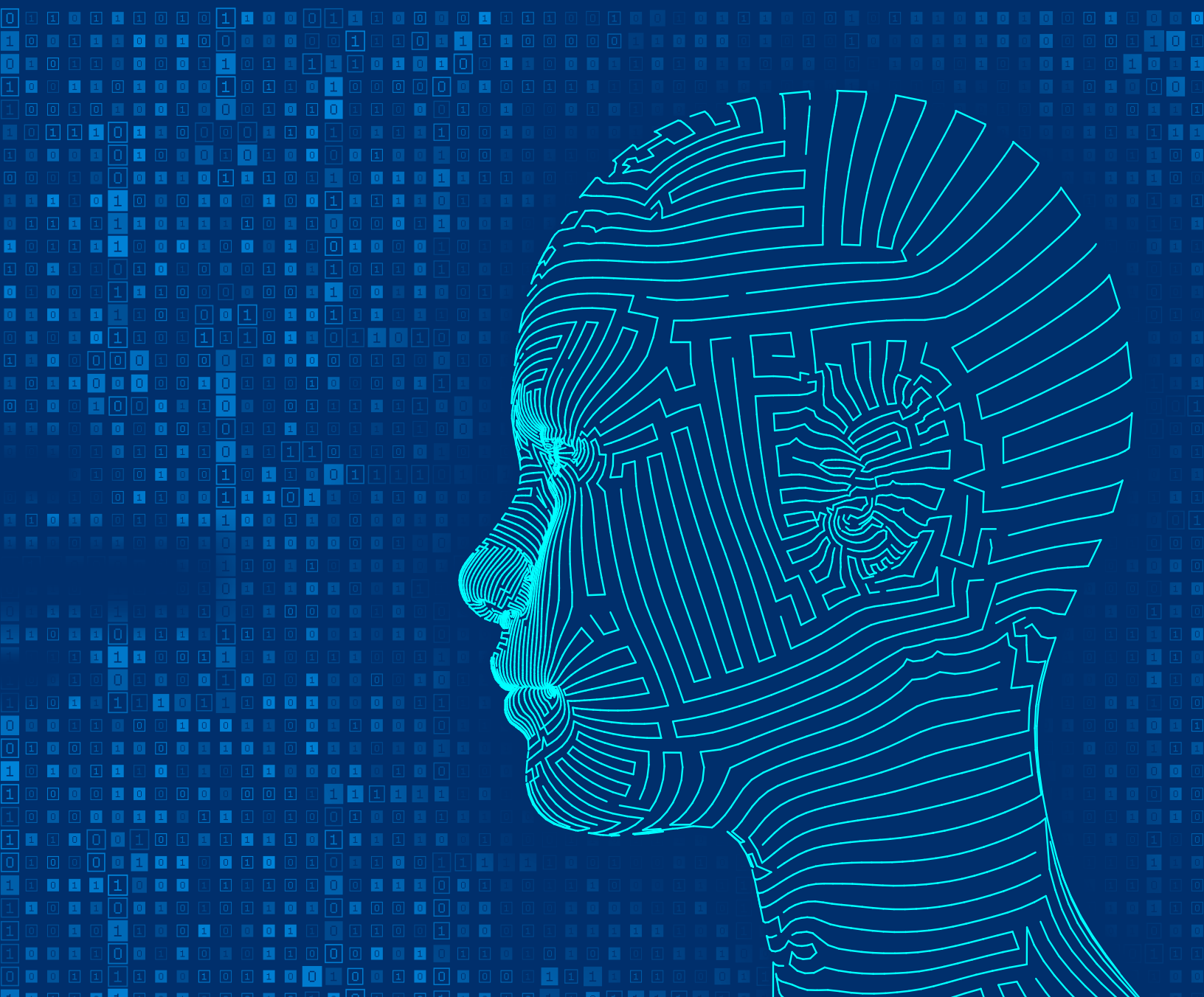
DAI Shaping a more livable world. | DigitalFrontiers SCALING DIGITAL DEVELOPMENT | ATHENA INFONOMICS | itad | WINDt Women in Digital Transformation | PIT POLICY LAB Public Interest Technology

This Guide is a resource for local-level policymakers across sectors seeking to understand the challenges, opportunities, and risks of AI adoption.

The Guide includes a definition of AI, a discussion of ethical issues related to AI, and proposals for how these issues can be addressed.

It also includes examples to illustrate AI-related ethical issues, approaches for addressing these issues, and relevant reflection questions. Specifically, this Guide uses the education sector as an example to explore elements of ethical AI, but the lessons are generalizable to other sectors as well.

Finally, the Guide includes a Glossary and an Annex with additional details of case studies highlighted in the document.



WHAT IS ARTIFICIAL INTELLIGENCE (AI)?

Artificial intelligence, or AI, is a concept referring to computer algorithms that solve problems using techniques associated with human intelligence: logical reasoning, knowledge representation and recall, language processing, and pattern recognition. AI is currently used to build a wide variety of applications—from customer service chatbots to complex earthquake or crime prediction programs.

WHY DO WE USE AI?

AI offers an exciting extension of many human capabilities such as observation, processing, and decision-making. The output and outcomes of AI systems are nearly instantaneous, offering humans powerful efficiencies that did not exist just a few years ago. The computing power and systems used for AI technologies far exceed human cognitive capabilities, allow for constant "machine learning" without human supervision, and include consideration of patterns that are typically impossible for humans to discern (e.g., the ability to identify an individual based on their gait without ever seeing their face). AI can also use dynamic nudging to create instant incentives for compliance (e.g., in the commercial context, a guided selection of benefits, personalized to the customer, designed to promote specific economic behavior).

HOW DOES AI WORK?

In an AI system, predefined algorithms follow a series of instructions to transform data into outputs that can be used for making decisions, either by the computer system itself or a human. Many AI algorithms learn directly from data, by identifying patterns and relationships, without rules-based instructions from humans—as is the case in traditional statistical programming.[1] The "black box" nature of AI systems refers to system inputs and operations which are not visible to the end-user or other parties, and sometimes even to the data scientists who build AI systems. Algorithms that can act independently without control or supervision from humans are considered autonomous.

For reflection

AI is dependent on the data it receives for decision-making. As with all computing systems, if the data used is biased or corrupted, so will be the outputs or recommended decisions by the AI system.

WHAT IS THE DIFFERENCE BETWEEN AI AND MACHINE LEARNING?

Machine learning (or ML) enables AI systems. While ML focuses on learning and prediction, AI applications often rely on the predictions from ML to create, plan, or have an impact in the real world. Automated decisions might be directly implemented (like in robotics) or suggested to a human decision-maker (in the form of product recommendations in an online shopping session).[2]

WHY DOES AI NEED DATA?

Large datasets, such as those generated by a state or national education system, enable AI. While AI systems have actually been around since the 1950s, AI has recently received increased attention due to its new capabilities, powered by larger quantities of available data and more powerful computing systems. For instance, data on population movements, the spread of epidemics, and climate change can significantly enhance the capabilities of today's AI systems to monitor emerging trends and provide evidence to advance on the UN Sustainable Development Goals.

For reflection

Access to sufficient, unbiased, good-quality data is a foundational driver of trustworthy and ethical AI.

1. Prasanna Lal Das (2022). Algorithms in Government: A Magic Formula or a Divisive Force? https://dial.global/research/algorithms-in-government-a-magic-formula-or-a-divisive-force/
2. USAID Digital Strategy (2020-2024). https://www.usaid.gov/digital-development/usaid-digital-strategy.

WHAT IS AI ETHICS?

Ethics is the science of proper behavior. Aristotle argued that ethics is the study of human relations in their most perfect form; he claimed that ethics is the basis for creating an ideal model of human interrelations, ensuring optimal communication between people and a reference point for creating a structure of moral consciousness. The practice of AI ethics is the consideration of moral problems related to the interaction of technology, humans, and society. Just as AI is evolving rapidly, so too are AI ethics considerations. Ultimately, the overall goal of AI ethics is to seek to create an optimal model of interrelations between humans and technology.

WHY DO WE USE AI IN THE EDUCATION SECTOR?

AI systems change the nature of educational systems through the ability to analyze large and complex datasets, automate processes, understand changes in student performance, create new ways of interaction between students and teachers, customize learning methods, and more. At the same time, the potential for careless implementation of AI creates serious ethical risks with long-term negative consequences, generates prejudice, can inhibit motivation, or even trigger social unrest.

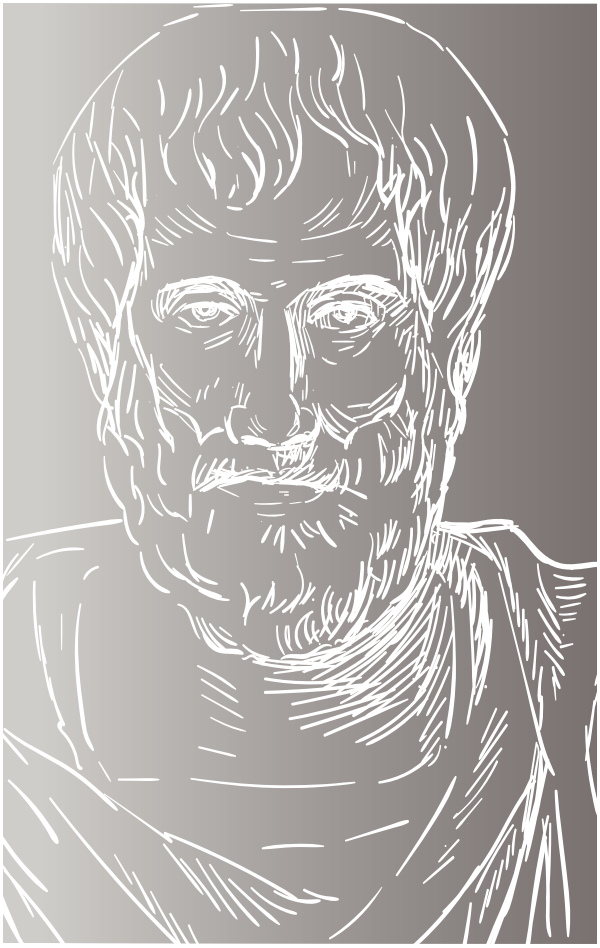
See more definitions of concepts related to AI ethics in the Glossary at the end of this Guide.

For reflection

The ethics of AI today is more about the right questions than the right answers.

For reflection

There is no common definition of AI ethics.



WHY AI ETHICS IS IMPORTANT

While AI can bring many benefits for human beings, it also comes with ethical considerations that cannot be ignored. As these systems are increasingly being used across sectors, AI can have significant effects on credit, employment, education, competition, and more. However, without ethics embedded in AI algorithms, it is hardly possible to guarantee that AI will not enable certain actors to cause more harm than good. With the wider adoption of AI in recent years, it has been used to sow distrust in public information and and has been held responsible for perpetuating discrimination in the delivery of services and unfavorably profiling segments of the population, raising many other moral concerns.

The responsibility of data scientists, teams of software developers, and other participants in the AI lifecycle3 rarely extends to AI ethics. Software engineers frequently pay more attention to how well a system or product is performing its intended function than to its social and ethical implications and longer-term ramifications. That is why it is important to engage different stakeholder groups, including a diverse range of members of civil society, to ensure ethical design and deployment of AI systems.

Case study

Imperfect AI solutions in education

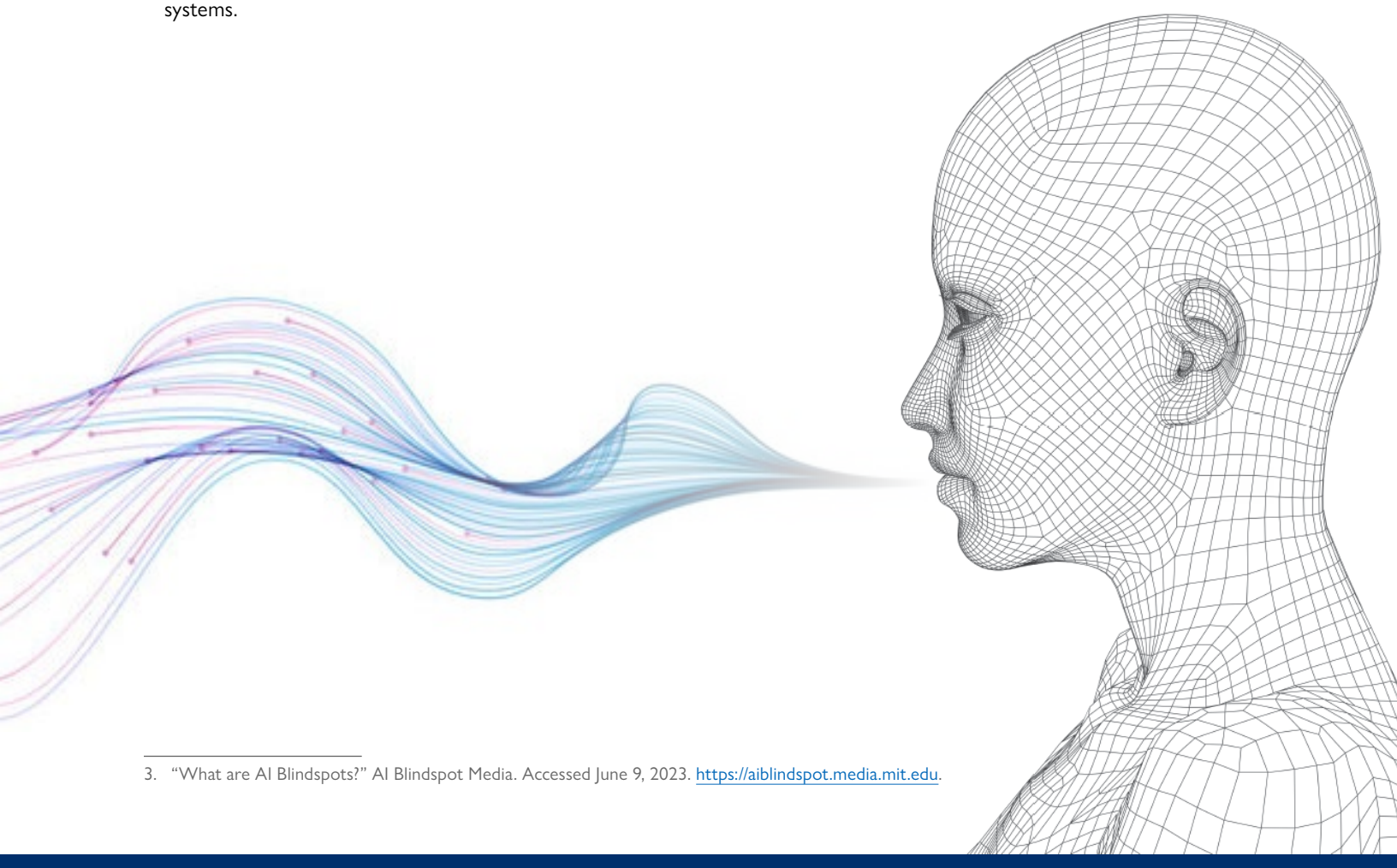
Country: UK

Year: 2020

What happened: The students enrollment algorithm favored students from private schools and affluent areas, leaving high-achievers from free, state-schools disproportionately affected.

Why it is important for the purposes of this Guide? Many policymakers focus only on positive effects that AI bring to the education sector, however, the identification and mitigation of AI's risks should be prioritized over implementation.

See details in the Annex.



3. "What are AI Blindspots?" AI Blindspot Media. Accessed June 9, 2023. https://aiblindspot.media.mit.edu.

WHAT ARE AI-RELATED MORAL PROBLEMS?

AI algorithms can be opaque and complex, and subject to error, bias, profiling, discrimination, and other unfair practices. This happens, in part, because algorithms are created by human AI engineers—and humans are not objective. This is also due to historical biases and prejudices in the datasets AI systems learn from. If AI engineers do not identify address data bias, AI systems may replicate them. Ethical issues appear in many aspects of AI adoption. Examples include the choices that self-driving cars are programmed to make when a crash is inevitable;[4] racial, gender, and age biases in facial recognition; biases for people with disabilities; and AI algorithms making discriminatory decisions in the insurance and banking sectors. There are recent cases in which people have blamed algorithms for injustices that negatively impact their lives, such as being denied bail by judges reliant on automated systems or their children being unfairly deprived of college admissions.[5]

For reflection

How would machines solve the "Trolly Dilemma"?

WHO IS RESPONSIBLE FOR AI OUTPUTS?

Humans curate datasets, create and serve as "role models" for algorithms, and consume the outputs of algorithms. Humans also provide an essential point of control for algorithms, either as testers or validators of algorithmic decisions. The societal question of who, exactly bears responsibility for errors in AI decision-making is a profound one: is the fallout from an AI-made error the fault of the developer of the AI algorithm, the owner of the smart device, the operator of the algorithm or AI-tool, the person who provided input data to the algorithm, or someone else entirely? The lack of certainty on this front may perpetuate and exacerbate distrust of AI systems, leading to long-lasting consequences for innovation and adoption of technology.

Case study

Flawed facial recognition systems

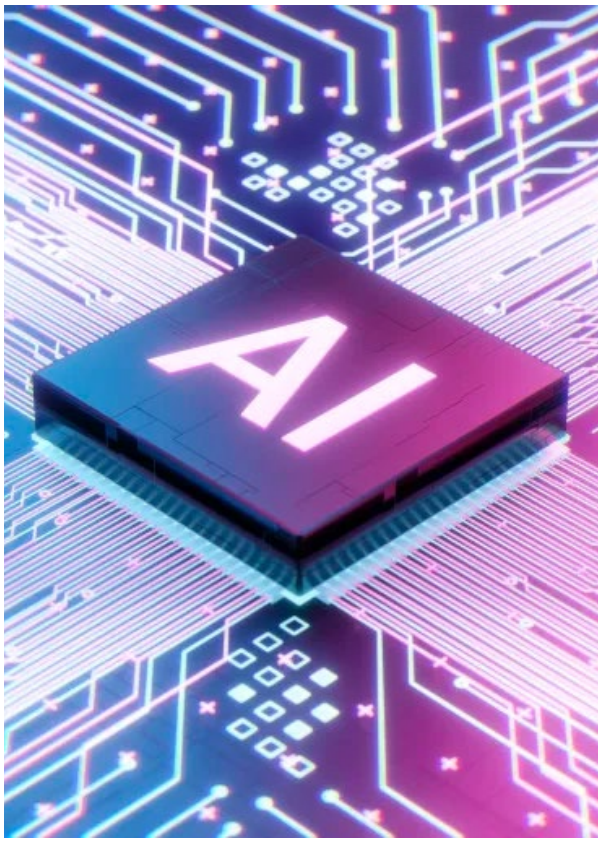
Country: USA

Year: 2018

What happened: Google facial recognition software had a bias against African Americans; image recognition algorithms in Google Photos were classifying African Americans as "gorillas."

Why it is important for the purposes of this Guide? Currently, numerous government entities consider the use of facial recognition systems, without considering these systems' potential risks and flaws, as illustrated by this use case.

See details in the Annex.



4. Orin Kerr. "A 'Trolley Problem' Dilemma for the FBI." The New York Times. January 27, 2016. https://www.nytimes.com/roomfordebate/2016/01/27/the-ethics-of-a-child-pornography-sting/a-trolley-problem-dilemma-for-the-fbi
5. "Racism In, Racism Out. A Primer on Algorithmic Racism." 2021. Public Citizen. https://www.citizen.org/article/algorithmic-racism/.

WHAT ARE SOME OF THE MORAL ISSUES SURROUNDING AI?

AI applications such as chatbots, criminal assessment systems, facial recognition systems have provided sexist or racist outputs because of biases inherent in input data. These biases are often found in open-source data, which is not necessarily representative of global populations. An AI algorithm draws conclusions from what it has studied; when deep learning systems are trained on open data, no one can control what exactly they learn. Using historical data may also create modern bias for AI tools as older datasets may not include data on women, people with disabilities, indigenous peoples, or other historically disenfranchised groups, many of whom are still on the "unfavorable" side of the digital divide. In some cases, an AI algorithm would likely consider these populations as outliers and create models that lead to further errors that perpetuate their disenfranchisements.

WHAT ARE THE RISKS?

Ethical problems in AI can lead to a variety of consequences with different levels of severity; these consequences include everything from increased inequality to extended litigation processes to resist social uprising. The profiling and biases of algorithms against a particular race, gender, or specific category of people can affect how education, healthcare, financial, and democratic systems work. AI can be also used maliciously, in any of these fields to fake data, steal passwords, and interfere with the work of other software and machines, thus undermining public trust in technology even more. These digital crimes put core human values such as personal privacy, data protection, fairness, and autonomy at risk.

For reflection

If an AI system generates an incorrect medical diagnosis that leads to the death of a patient, who is responsible?

Case study

AI against people with disabilities

Country: Global

What happened: Persons with disabilities may be interpreted as outliers by AI applications that may mimic the direct and indirect discrimination they face in society.

Why it is important for the purposes of this Guide? When considering the development and adoption of an AI system at the local level, it is important to identify if there is representative data that includes people with disabilities, for example, to avoid their exclusion from or discrimination within the results or recommendations of such a system.

See details in the Annex.

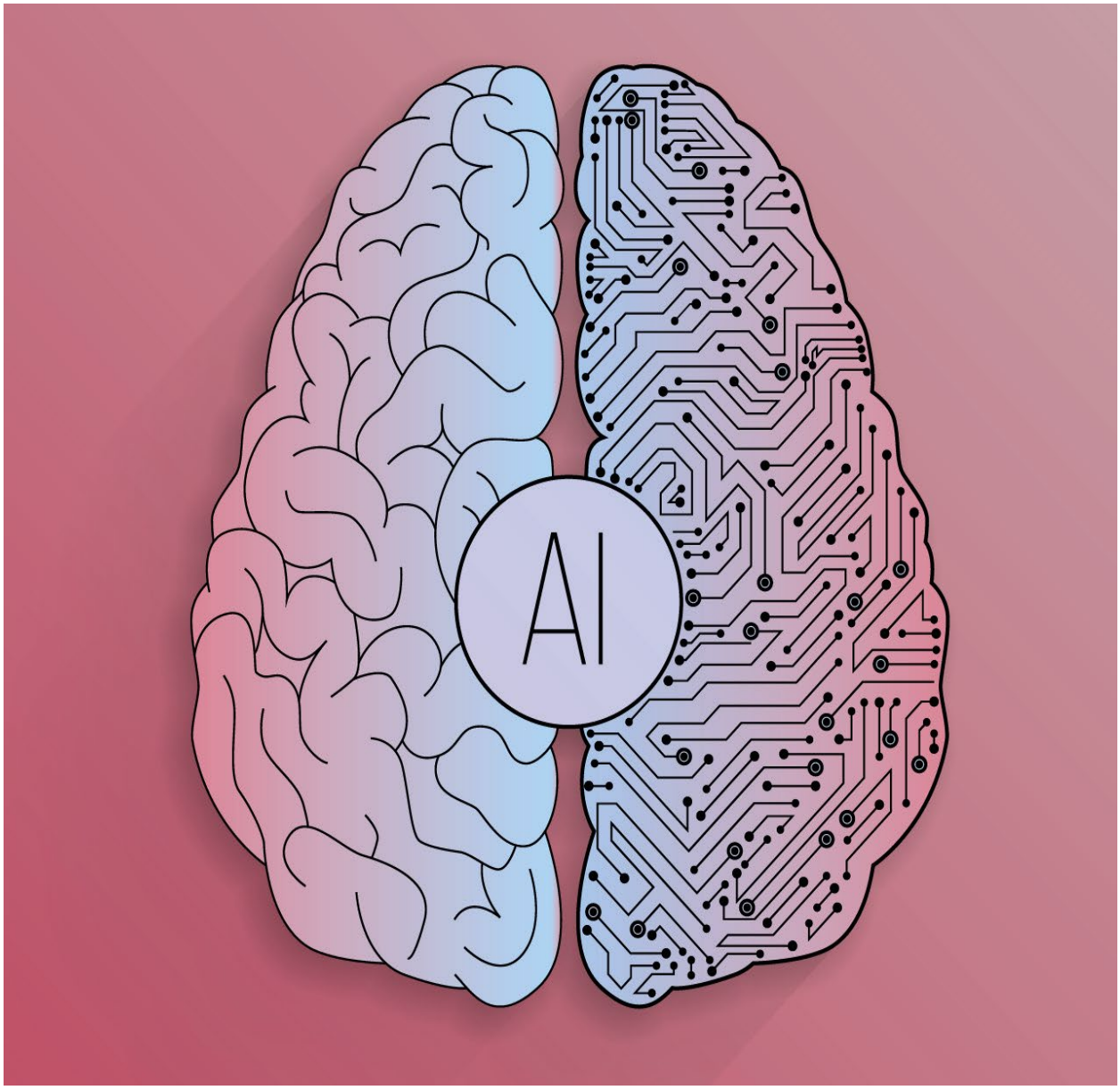
6. AI Decolonial Manyfesto." Mayfesto.ai. Accessed June 9, 2023. https://manyfesto.ai.
7. "Principles of Māori Data Sovereignty." 2018. Te Mana Raraunga. https://cdn.auckland.ac.nz/assets/psych/about/our-research/documents/TMR+M%C4%81ori+Data+Sovereignty+Principles+Oct+2018.pdf

WHO SHOULD BE INTERESTED IN AI ETHICS?

Everyone. Citizens, businesses, governments, and academia are all users or adopters of AI applications and technologies. They all face, or will face, AI-related ethical challenges that will impact their lives or livelihoods in some way. And they all have a perspective to add to our society's conversation surrounding ethical AI use.

DO AI SYSTEMS HAVE "GLOBAL" BIASES?

AI systems development, data collection, and standards creation are happening in the Global North, where most AI research institutions and big tech companies are located, while most countries of the Global South still experience data poverty and are still forming reliable and robust digital infrastructures. Thus, there are concerns about whether Western conceptions of "fairness" and equality should be considered universal,[6] and if they should be applied in the same manner in developing countries as they are in advanced economies. Many parties in developing nations have openly questioned the presence of Western "values" in AI systems and wonder whether alternative beliefs held in developing nations could inform ethical AI systems that could be accepted globally. Other concerns, related to maintaining the data privacy of global indigenous peoples in the face of new AI technologies, have been raised as well. One example is Maori Principles for Data Sovereignty.[7]



HOW TO APPLY AI ETHICS

Governments around the globe are considering both the opportunities and risks of AI deployment. Certain entities have drafted ethical guidelines for use of AI and other automated systems, and while many of these guides are produced by institutions or governmental offices without explicit legal or legislative powers, such guides create a baseline of rules, begin to establish standards, and build awareness for prominent government actors. Further, private sector organizations, academic institutions, and non-government organizations[8] have also created their own sets of AI ethical principles or guidelines for a wider audience. Below, you will find several key suggestions that appear in one form or another in many these guides that can help foster the ethical use of AI:

1. CREATE AWARENESS AND INITIATE DIALOGUE

AI ethics is equally important for government, businesses, and individuals to understand; however, most people do not have a full understanding of it. Thus, a good starting point during AI implementation is to launch awareness activities that bring these stakeholders to the table. It is important to ensure that these activities catalyze wide-ranging engagement and discussion on AI ethics issues and how they impact national, sub-national, and municipal agencies; small- and medium-sized enterprises (SMEs) and larger businesses; professional associations; and citizens.

Case study

AI against people with disabilities

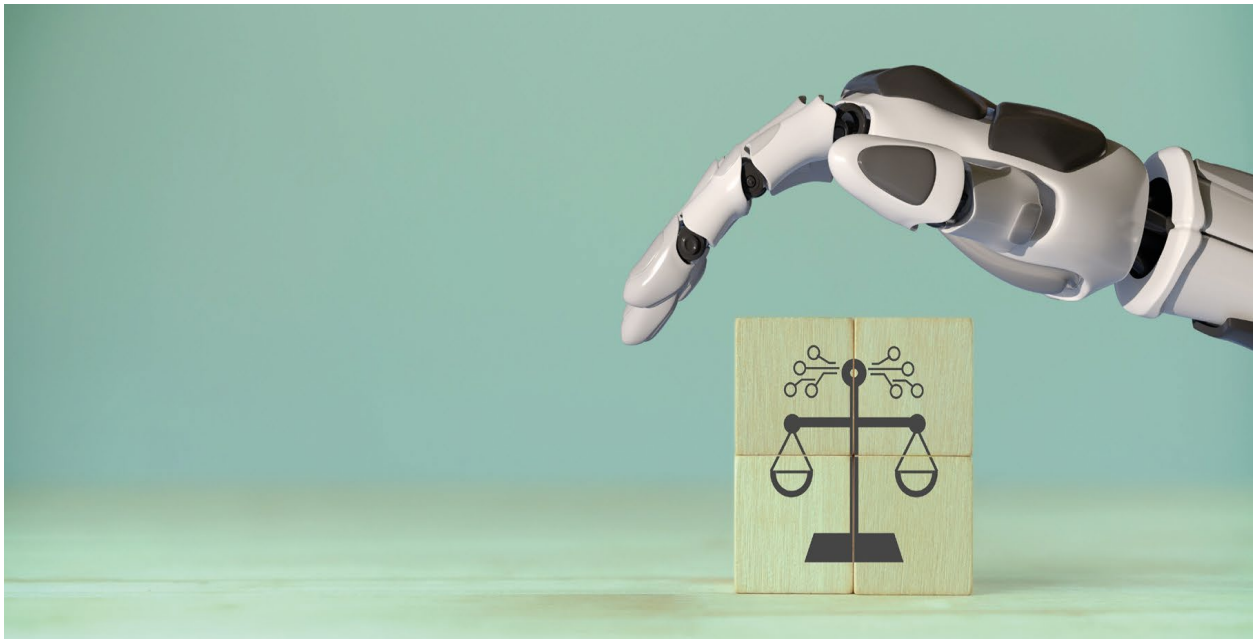
Country: Germany

Year: 2018

What happened: The Data Ethics Commission was established to produce ethical benchmarks, guidelines, and recommendations for the development and use of AI for the government.

Why it is important for the purposes of this Guide? This good practice can be contextualized by local governments that seek to have specialized work groups for responsible use of AI on this matter.

See details in the Annex.



8. "AI Ethics for Latin America." 2022. C Minds. https://www.cminds.co/booklet.

2. ADOPT GUIDING PRINCIPLES

Many countries, cities, and organizations have developed their own principles and guidelines, codes of ethics, and self-assessment toolkits on AI for public officials and business organizations.[9] At the heart of each independent document lies a universal list of guiding principles that do not bear any legislative power but help to shape ethical application of AI for all. The general list comprises the following five principles:

Transparency

AI uses open, understandable, auditable, and documented use of any data with clear and up-to-date security measures.

Non-maleficence and trust

AI does no harm or inflicts the least possible harm to reach an outcome. Continuous operator readiness to interfere in AI systems is ensured.

Equitability and fairness

AI systems and solutions subordinate to human- defined rules and laws, thus ensuring human rights above all.

Autonomy and responsibility

The AI system discloses who is responsible for AI works and solutions. At the same time, it recognizes that humans willingly give up some of decision-making power to AI.

Explainability

To ensure accountability, AI systems and their decision-making processes need to be explainable to a human being at a basic and acceptable level.

There are other AI guiding principles adopted by different organizations and that are fixed in internationally recognized documents, such as the Montreal Declaration for Responsible AI[10] or OECD's principles,[11] which are worth getting acquainted with in detail.

Case study

Canada sets up guiding principles

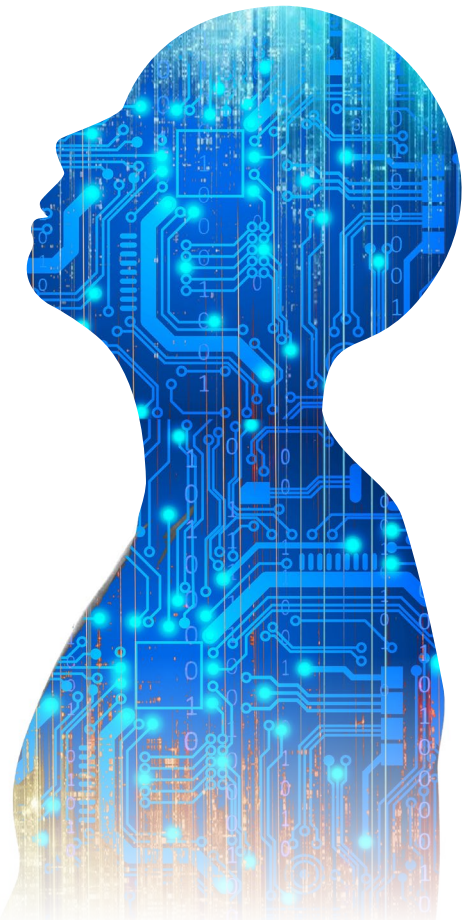
Country: Canada

Year: 2018

What happened: The Treasury Board developed a set of guiding principles to help government officials explore AI in a way that is "governed by clear values, ethics, and laws.

Why it is important for the purposes of this Guide? Aspects and lessons learned from the Canadian experience can be considered in the design of guiding principles aligned to the local reality of state and municipal governments.

See details in the Annex.



9. See, for instance, the UN Moratorium on use of AI that threatens human rights, or the https://fairlac.iadb.org/en/fair-lac-box (2022).
10. "The Montréal Declaration for Responsible AI Development." 2018. Université de Montréal. https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_Artificial_Intelligence_2018.pdf.
11. "The OECD AI Principles." 2019. OECD.AI Policy Observatory. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

3. DEVELOP A LEGAL FRAMEWORK FOR AI ETHICS

Most AI ethics initiatives developed today have inspired AI guidelines and frameworks that are recommended but not compulsory. Developing ethical standards for data processing using AI can be an important step towards setting up a legal foundation for ethical AI. The standard can be in the form of methodological recommendations for developing ethical codes for organizations participating in the development and introduction of AI technologies.

Case study

Open explanations of AI from Singapore

Country: Singapore **Year**: 2020

What happened: An AI startup disclosed the exact parameters used in developing its AI model to its clients in the healthcare sector. The startup made a conscious decision to declare the use of AI and in its analysis and prediction.

Why it is important for the purposes of this Guide? This example sets a precedent for a successful business case where there is a balance between transparency in the use of AI and privacy of user data.

See details in the Annex.

4. DRAFT SECTOR-SPECIFIC GUIDELINES FOR THE USE OF AI

AI ethics varies when it comes to AI implementation in specific sectors. Expecting actors in sectors like healthcare, e-commerce, and education to simply adhere to common principles of fairness and justice when using AI will generally not be a sufficient approach; instead, it is pivotal for guidelines for AI use in these sectors will need to provide a detailed framework relevant to each sector to guide their deployment of AI technologies. Sectoral agreements on such frameworks can be beneficial for the ubiquitous application of AI.

Case study

Resource for UK public authorities

Country: UK **Year**: 2020

What happened: The government developed an AI guide for public authorities to better understand how AI works and how it can be applied in the public sector.

Why it is important for the purposes of this Guide? The UK is a leading country in developing guidelines for the responsible and ethical use of AI. This exercise represents a good practice that groups guiding principles on the matter.

See details in the Annex.

5. ENCOURAGE TRANSPARENCY AND DISCUSSION OF NOVEL DATA USES WITH ETHICAL IMPLICATIONS AS THEY EMERGE

The data economy is a highly dynamic field, and the adoption of laws, norms, and standards often does not keep pace with the introduction and use of technologies—especially AI technologies. Permanent and practical discussions around emerging ethical implications of AI are important if a sustainable digital and data-driven transformation is to be achieved. Close collaboration with civil society could help monitor these rapid technological changes and ensure continued transparency regarding AI.

6. ENGAGE WITH THE DEVELOPMENT OF INTERNATIONAL PRINCIPLES OF AI ETHICS

Countries such as the United Kingdom, Canada, Germany, Japan, Argentina, the United Arab Emirates, along with international organizations such as the European Commission, the Organization for Economic Co-operation and Development (OECD), the Global System for Mobile Communications Association (GSMA), and many others actively contribute to the AI Ethics agenda by issuing codes of ethics and guidelines and organizing consortiums such as the Global Partnership for AI (GPAI) to work together on facilitating ethical adoption of AI.[12]

Case study

Resource for UK public authorities

Country: Global **Year**: 2019

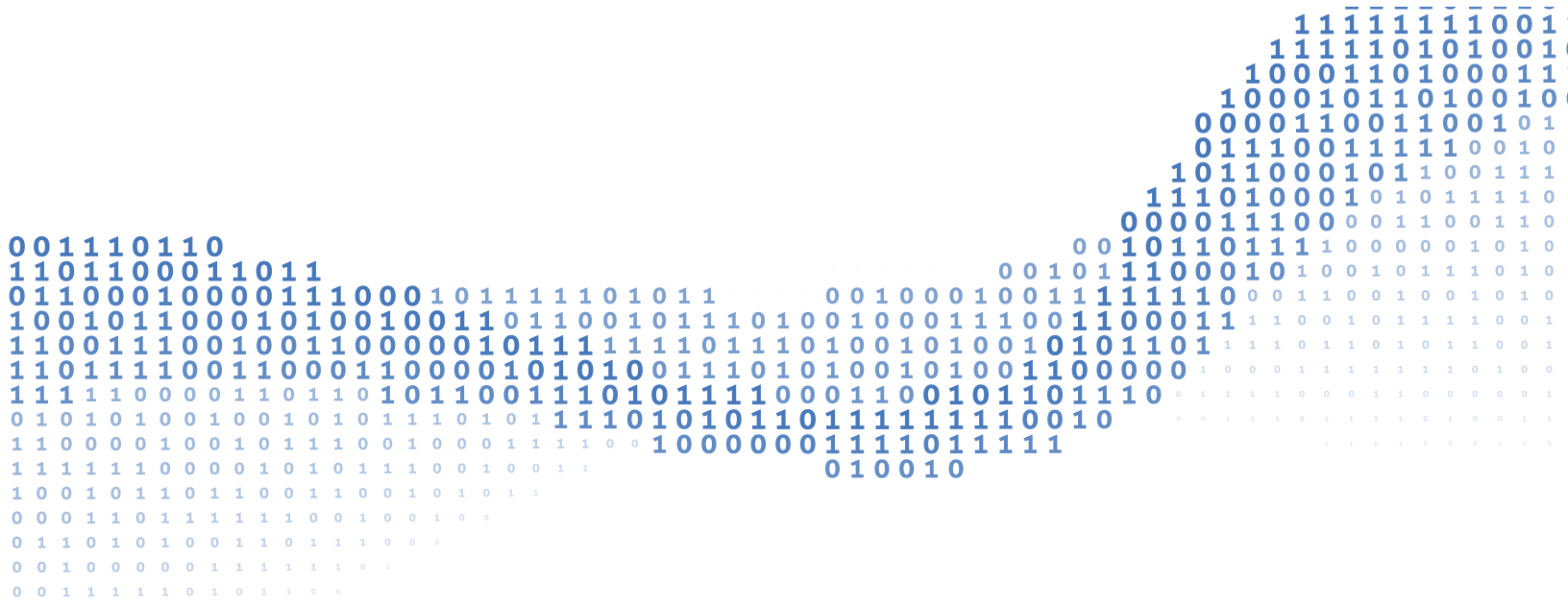
What happened: OECD published its AI Principles, which have been accepted by all 37 OECD countries and seven non-member partners, including Mexico. These principles also formed the basis of G20 AI Principles.

Why it is important for the purposes of this Guide? It is important that the local government exploring the potential use of AI takes into consideration whether their country has adhered to the OECD AI Principles in order to take them as a starting point for any public policy development.

See details in the Annex.

7. UTILIZE PRIVATE SECTOR INITIATIVES FOR PUBLIC GOOD

The private sector is also deploying publicly available AI tools that help to mitigate AI risks. Examples include the AI Fairness 360 tool by IBM, Google's People + AI Research (PAIR), the Aequitas Bias and Audit Toolkit from Carnegie Mellon University, Fairlearn, Datasheets for Datasets from Microsoft, and Tensorflow Model Cards. These tools have profound potential applications for AI ethics; for instance, the IBM AI Fairness tool is a comprehensive open-source toolkit of metrics to check for unwanted bias in datasets and machine learning models, and its state-of-the-art algorithms to mitigate such bias are applied by many practitioners. USAID also supports these endeavors. In 2021, the Agency introduced Managing Machine Learning Projects in International Development, a practical toolkit for those in governments or public sectors working with work with data scientists to develop AI.



12. "The Global Partnership on Artificial Intelligence." GPAI.ai. Accessed June 9, 2023. https://gpai.ai/

GLOSSARY

Definitions below are derived from multiple USAID studies,[13] unless specified.

Adoption: Changes that happen when people or institutions begin to use a new technology and incorporate it into their existing routines or processes. For example, people who use a mobile-money account to receive remittances and pay bills would be considered "adopters," while those who make a one-time withdrawal to empty a cash-transfer account would not.

Algorithm: A step-by-step procedure to turn any given inputs into useful outputs. A computer algorithm follows a series of instructions to transform inputs (data) into outputs that can be used for making decisions, either by the computer system or a human.

Internet of Things (IoT): Refers to connected devices and machines that gather data, connect it with intelligent analytics, and adapt their behavior/responses based on the information in the communication network. Smartphones are IoT devices.

Cybersecurity: The prevention of damage to, protection of, and restoration of computers, electronic communications systems, electronic communications services, wire communication, and electronic communication, including information contained therein, to ensure its availability, integrity, authentication, confidentiality, and non-repudiation.

Cyber Hygiene: The practices and steps that users of computers and other devices take to maintain system health and improve online security. These practices are often part of a routine to ensure the safety of identity and other details that could be stolen or corrupted.[14]

Digital Literacy: The ability to access, manage, understand, integrate, communicate, evaluate, and create information safely and appropriately through digital technologies for employment, decent jobs, and entrepreneurship.[15]

Machine Learning: The statistical process of deriving a rule or pattern from a large body of data to predict future data.

Open Data: Refers to data made freely available and deliberately stored in an easily read data format, particularly by other computers, and thereby repurposed.

Data Privacy: The right of an individual or group to maintain control over, and the confidentiality of, information about themselves, especially when that intrusion results from undue or illegal gathering and use of data about that individual or group.

Data Protection: The practice of ensuring the protection of data from unauthorized access, use, disclosure, disruption, modification, or destruction, to provide confidentiality, integrity, and availability.

Digital Economy: The use of digital and Internet infrastructure by individuals, businesses, and government to interact with each other, engage in economic activity, and access both digital and non-digital goods and services. A diverse array of technologies and platforms facilitate activity in the digital economy; however, much activity relies in some measure on the Internet, mobile phones, digital data, and digital payments.

Digital Infrastructure: The foundational components that enable digital technologies and services. Examples of digital infrastructure include fiber-optic cables, cell towers, satellites, data centers, software platforms, and end-user devices.

Digital Literacy: The ability to "access, manage, understand, integrate, communicate, evaluate, and create information safely and appropriately through digital devices and networked technologies for participation in economic and social life."4 This may include competencies that are variously referred to as computer literacy, information and communications technology (ICT) literacy, information literacy, and media literacy.

Platform: A group of technologies used as a base upon which other technologies can be built or applications and services run. For example, the Internet is a platform that enables web applications and services.

13. "USAID DECA Toolkit." U.S. Agency for International Development. 2022. https://www.usaid.gov/digital-development/deca-toolkit.
"USAID Digital Strategy." U.S. Agency for International Development. 2018. https://www.usaid.gov/digital-development/usaid-digital-strategy.
"USAID Managing Machine Learning Projects in International Development: A Practical Guide." U.S. Agency for International Development. n.d. https://www.usaid.gov/digital-development/managing-machine-learning-projects.

14. Chris Brook, "What is Cyber Hygiene? A Definition of Cyber Hygiene, Benefits, Best Practices, and More." Fortra. May 6, 2023. https://www.digitalguardian.com/blog/what-cyber-hygiene-definition-cyber-hygiene-benefits-best-practices-and-more.

15. "The Digital Literacy Global Framework (DLGF). 2018. The United Nations Educational, Scientific and Cultural Organization. https://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf.

16. USAID Digital Literacy Primer. U.S. Agency for International Development. 2022. https://www.usaid.gov/digital-development/digital-literacy-primer.

ANNEX: ETHICAL AI CASES IN DETAIL

Examples of cases related to the ethics of AI, representing both good practices and negative impacts.

AI DOWNGRADES STUDENTS

In the UK, the students enrollment algorithm favored students from private schools and affluent areas, leaving high-achievers from free, state-schools disproportionately affected. Subsequent reviews suggested that the algorithms might have been biased (reinforcing prejudices in historical data, plus favoring smaller schools). Additionally, many students had their university places revoked after the algorithm downgraded students who temporarily could not sit for exams due to the COVID-19 pandemic. Students took to the streets and the courts for redress, forcing the government to retract the lower grades. Source.

FACING PRIVACY AND SURVEILLANCE ISSUES WHILE EMBEDDING AI IN EDUCATION

Artificial intelligence can help students get useful feedback faster and, among other things, reduce the burden on teachers. Artificial intelligence systems can track how users interacts with digital tools; the resulting experience provides a personalized experience. In education, this may include systems that identify strengths and weaknesses and patterns in student performance. While teachers do this to some extent in their teaching, monitoring and tracking online conversations and student actions may also limit student participation, as students feel uncomfortable knowing their activities and responses are being monitored and analyzed by AI. Source.

FLAWED FACIAL RECOGNITION SYSTEMS

Google is one of the leaders in AI, however, its facial recognition software has shown bias against African Americans. In several stated cases, image recognition algorithms in Google Photos were classifying African Americans as "gorillas." Google has not yet found a solution for this error, instead implementing a workaround in which it erased gorillas and several other primates from the platform's lexicon to prevent further misclassification. Source.

LACK OF GENDER-DISAGGREGATED DATA LEADS TO FATAL OUTCOMES

Crash-test dummies were first introduced in the 1950s, and for decades they were modeled on data for the 50th-percentile male. The most commonly used dummy is five foot nine inches (175cm) tall and weighs 167 pounds (75.74kg) – measurements which are significantly larger than an average woman. This male-skewed data is factored into automotive design and is one reason why women are 17 percent more likely to be killed and 47 percent more likely to be injured in car crashes than men. A lack of gender-disaggregated data also affects modern digital tools. A recent study of common fitness monitors found that they underestimated steps women take during housework by up to 74 percent and underestimated calories burned during housework by as much as 34 percent. Further, women are significantly excluded from medical research because the findings are not gender disaggregated. This lack of disaggregated data means that half of the population is misrepresented or does not have equal access to public or private services and technologies. Source.

AMAZON AI AGAINST WOMEN

Amazon abandoned its AI hiring program because of its bias against women. The algorithm began training on the resumes of the candidates for jobs posted over the previous ten years. Because most of the applicants were men, it developed a bias to prefer men and penalized features associated with women. The program failed to remove the gender bias as it was profoundly embedded in the datasets. Source.

POLICE AI ACTS AGAINST BLACK PEOPLE

A study of AI tools that U.S. authorities use to determine the likelihood that a criminal reoffends found that algorithms produced different results for Black and white people under the same conditions, with Black people being defined as more likely to reoffend. This controversial use of AI for "predictive" policing, has led to sharp criticism of law enforcement's use of AI. Source.

RACIST AI BOT ON TWITTER

In 2016, Microsoft launched an AI bot on Twitter that could interact and learn from users of the social media platform. However, it exhibited racist and sexist behaviors a few hours after learning from open data on Twitter. Microsoft shut it down less than a day after its release. Source.

AI-ENABLED TRANSLATION SERVICES ARE BIASED TOWARDS MALES

Google Translate systematically changes the gender of translations in a way that favors males. Stereotypes sneak into translations because Google optimizes translations for English. Experiments showed that, in many cases, Google changes the gender of the word; for instance, the German phrase "vier Historikerinnen und Historiker" (four male and female historians) is rendered as "cuatro historiadores" (four male historians) in Spanish, meaning that female historians were removed from the text. Google Translate provided similar results in Italian, French and Polish. Source.

AI AGAINST PEOPLE WITH DISABILITIES

There are examples of unintentional discrimination against persons with disabilities by AI applications. Persons with disabilities may be interpreted as outliers by an AI application, mimicking the direct and indirect discrimination these persons face in society. For example, AI systems programmed on past employee data may interpret a disclosed disability as a negative characteristic if past applicants with disabilities were frequently screened out at early stages. Source.

AI "PREDICTED" TEEN PREGNANCY

In 2018, the Ministry of Early Childhood in the Argentinian province of Salta partnered with Microsoft to pilot an algorithmic system to predict teenage pregnancy. They called it the Technology Platform for Social Intervention. The goal of the pilot was to forecast which females from low-income areas would become pregnant within the next five years. The implications upon women and girls of being declared "predestined" for motherhood, or how knowing this information would then help prevent adolescent conception, were not acknowledged or publicly discussed by the ministry or Microsoft. The system was based on analyzing data—including age, ethnicity, country of origin, disability, and whether the subject's home had hot water in the bathroom—from 200,000 residents in the city of Salta, included 12,000 women and girls between the ages of 10 and 19. The Technology Platform for Social Intervention was never the subject of a formal review, and its effects on girls and women have not been examined since due to the absence of national AI legislation. No formal information about the platform's accuracy or results has been released since. However, it has since been discovered that the system's database only contained information on racial and socioeconomic groups and did not have information on access to sex education or contraception, which are widely acknowledged by public health organizations as the most effective methods for lowering the rate of teen pregnancy. Source.

AI Ethics Solutions

GERMAN DATA ETHICS COMMISSION

In Germany, a Data Ethics Commission was established to produce ethical benchmarks, guidelines, and recommendations for the development and use of AI for the government. The result of their work was a publication of ethical guidelines with a set of Specific Recommendations for Action, with the intention of "protecting the individual, preserving social cohesion, and safeguarding and promoting prosperity in the information age." Notably, in Germany, many private organizations establish their own overarching AI guidelines. For instance, Deutsche Telekom established Guidelines for Artificial Intelligence that describe how AI at Deutsche Telekom should be used and how AI-based products should be developed.

GUIDING PRINCIPLES FROM CANADA

Canada is exemplary in its deployment of tools for public officials that help them explore AI in ways that are "governed by clear values, ethics and laws." The Canadian approach is comprehensive: it provides Guiding Principles to ensure ethical use of AI, a list of businesses looking to sell AI solutions to the government, and an algorithm impact. The latter helps government bodies assess the potential risks of deploying an automated decision-making system. It is presented in the form of an 80-point questionnaire related to business process, data, system design, algorithm, and system design decisions. The results provided by the assessment inform the body around the potential impact of the proposed AI and provide information about applicable requirements.

AI LAWS IN THE USA

The National Artificial Intelligence Initiative was adopted in 2020.[17] However, this initiative has not defined bias or mentioned gender bias as of this review. A new Blueprint for an AI Bill of Rights from the Biden-Harris Administration has introduced five protections necessary for individuals in the AI age. Specifically, the Algorithmic Discrimination Protection declares, "You should not face discrimination by algorithms, and systems should be used and designed equitably."[18] In October 2022, the Law on Artificial Intelligence Training for the Acquisition Workforce Act (also known as the "AI Training Act")[19] was adopted to provide an AI training program for the acquired workforces of executive agencies. The purpose of the program is to ensure the public workforce has knowledge of the capabilities and risks associated with AI. Meanwhile, several federal and state government agencies and private sector actors have launched initiatives to prevent algorithmic bias. An industry initiative led by the Data & Trust Alliance has developed Algorithmic Bias Safeguards for the Workforce, a structured questionnaire that businesses can use for procuring software to evaluate workers.

OPEN EXPLANATIONS OF AI FROM SINGAPORE

UCARE.AI, a Singapore-based startup, offers an AI-powered "cost predictor" product on its platform that works with hospitals to deliver accurate estimations of hospital bills to patients. To build greater confidence and trust in the use of AI, the company was mindful to be transparent in its use of AI with various stakeholders. UCARE.AI not only disclosed the exact parameters used in developing the AI model to its clients, but also provided detailed explanations on all algorithms that had any impact on operations, revenue, or customer base. Realizing that the accuracy of bill projection is highly regarded by hospitals and patients, UCARE.AI made a conscious decision to declare the use of AI in its analysis and prediction of bill amounts to client data managers and its patients. Source.

AI GUIDE FOR PUBLIC AUTHORITIES IN THE UK

The UK Government developed a guide to using AI in the public sector to better understand how AI works and how it can be applied in the public sector. A final illustrative document from the government is the Guidelines for AI Procurement that provide a set of principles on how to vet and buy AI technology, as well as insights on tackling challenges that may arise during procurement.

AI ETHICS—A JOINT EFFORT

Recognizing that issues relevant to AI transcend borders, countries are increasingly adopting regional approaches to AI, including coordinated efforts in the European Union and the African Union, among Nordic-Baltic states and Arab nations, and within the G7 and the G20. The OECD has also strengthened its AI-related efforts in recent years, spearheaded by the OECD.AI Policy Observatory. Indeed, the OECD AI Principles adopted in 2019 are the first intergovernmental standards on AI. The OECD created its guidelines that provide a list of overarching principles and policy recommendations with an aim to guide governments, organizations, and individuals in designing and running AI systems in a way that puts people's best interests first and ensuring that designers and operators are held accountable for their proper functioning. Source.

17. "National Artificial Intelligence Initiative Act of 2020." 2020. Congress.gov. https://www.congress.gov/bill/116th-congress/house-bill/6216.
18. "Algorithmic Discrimination Protections." 2020. The White House https://www.whitehouse.gov/ostp/ai-bill-of-rights/algorithmic-discrimination-protections-2/.
19. "Artificial Intelligence Training for the Acquisition Workforce Act or the AI Training Act." 2022. Congress.gov. https://www.congress.gov/bill/117th-congress/senate-bill/2551.

USAID
FROM THE AMERICAN PEOPLE

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, DC 20523
Tel: (202) 712-0000
www.usaid.gov