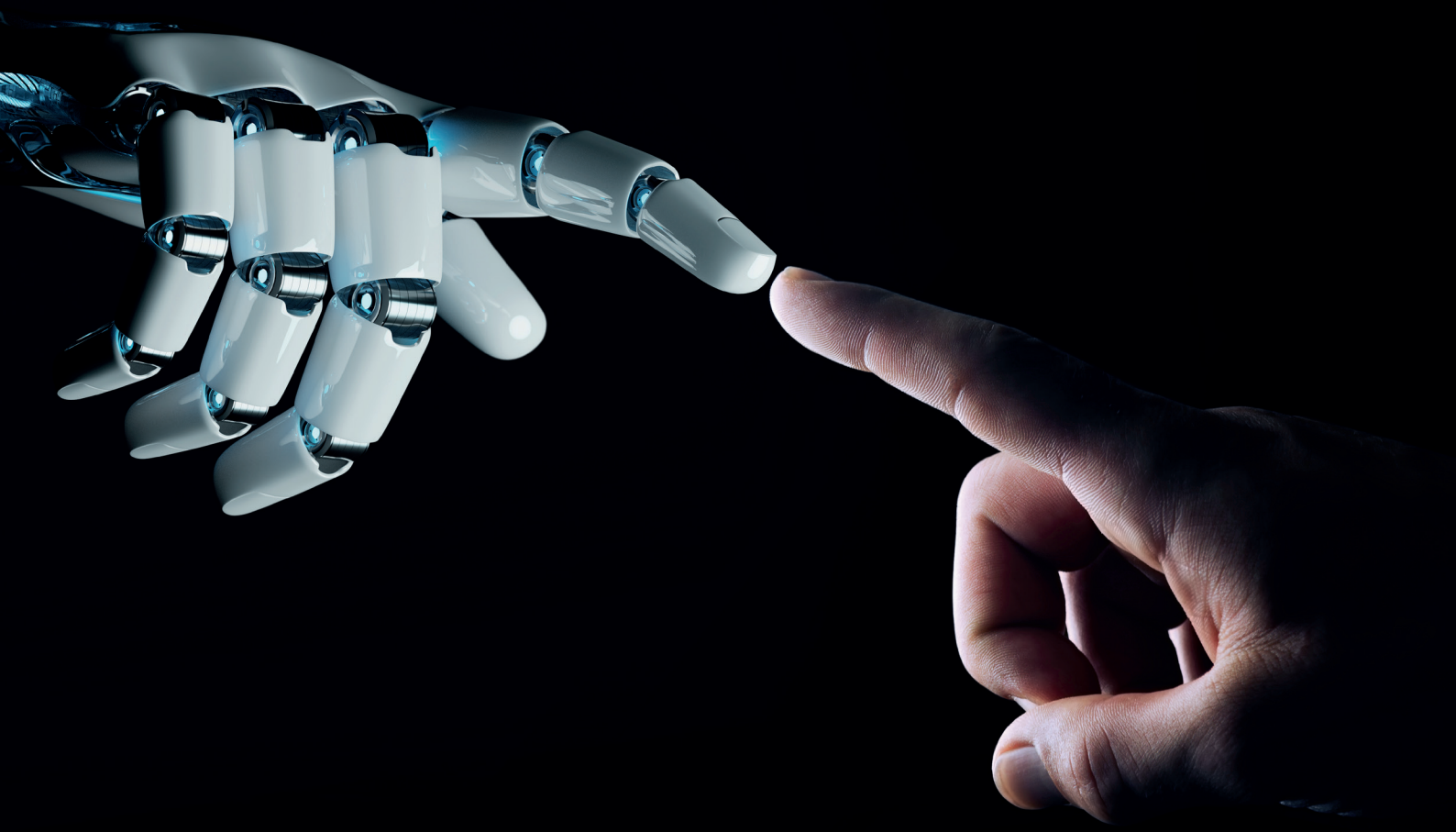
USAID
FROM THE AMERICAN PEOPLE



ARTIFICIAL INTELLIGENCE ETHICS GUIDE

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

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

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

DISCLAIMER This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of DAI and do not necessarily reflect the views of USAID or the United States Government. This publication/report/guide was produced under DAI's Digital Frontiers Project (Cooperative Agreement AID-OAA-A-17-00033) at the request of USAID

CONTENTS

WHAT IS ARTIFICIAL INTELLIGENCE?

Artificial intelligence (or AI) is a concept referring to computer algorithms that solve problems using techniques associated with human intelligence: logical reasoning, knowledge representation, language processing, and pattern recognition. AI is used to build a wide variety of applications - from customer service chat-bots 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 offer efficiencies and efficacy to humans not otherwise possible. The computing power and systems used for AI technologies far exceed human cognitive capabilities, allow for constant learning without human supervision and include consideration of patterns that are typically impossible for humans to discern (e.g., the ability to identify individuals based on their gait without ever seeing their face). AI can also use dynamic nudging to create instant incentives for compliance (e.g., a guided selection of benefits designed to promote specific economic behavior).

HOW DOES AI WORK?

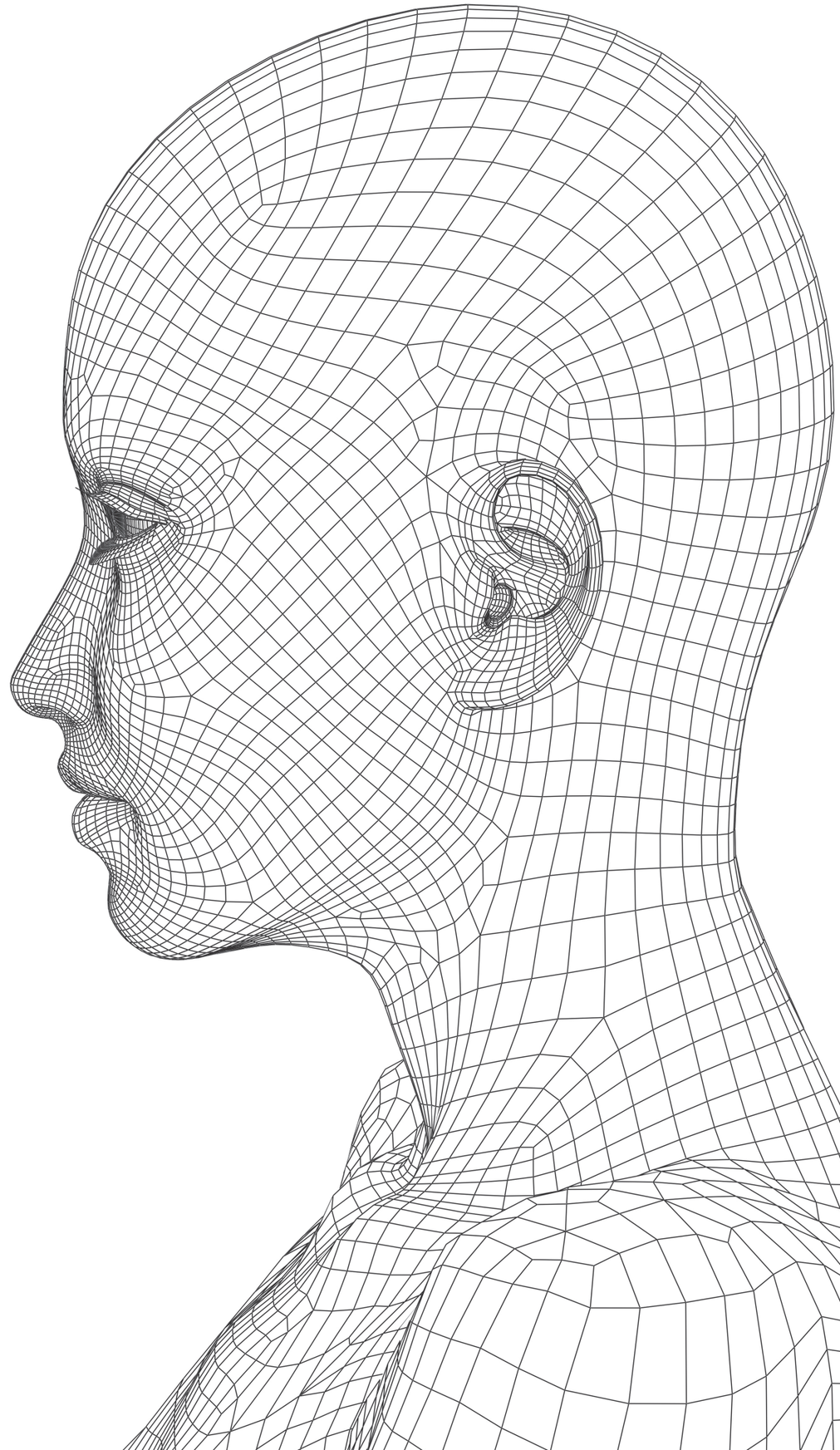
In an AI system, pre-defined algorithms follow a series of instructions to transform data into outputs that can be used for making decisions, either by the computer system itself or a human. Many AI algorithms learn directly from data, by identifying patterns and relationships, without rules-based instructions from humans as is the case in traditional statistics.[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.

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 (like product recommendations in online shopping).[2]

For reflection

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



1 Prasanna Lal Das (2022). Algorithms in Government: A Magic Formula or a Divisive Force?

2 USAID Digital Strategy (2020-2024).

WHY DOES AI NEED DATA?

Large datasets, like those generated by a state or national education system, enables AI. While AI systems have been around since the 1950's, AI has received increasing attention because of increased quantities of available data and computing powers. For instance, data on population movements, the spread of epidemics, and climate change can significantly enhance the capabilities of AI to monitor emerging trends and provide evidence to advance on the UN Sustainable Development Goals.

WHAT ARE 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 optimal 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 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 data sets, 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

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

For reflection

There is no common definition of AI ethics.

For reflection

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

WHY AI ETHICS IS IMPORTANT

While AI can bring many benefits for human beings, there are some 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 algorithms, it is hardly possible to guarantee that they do not 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 government machinery, 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 lifecycle[3] 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 ethical social and implications. That is why it is important to engage different stakeholder groups such as 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, while identifying and mitigating risks should be prioritized first.

See details in the Annex.

3 What are AI Blindspots

WHAT ARE AI-RELATED MORAL PROBLEMS?

AI algorithms can be opaque, complex, and subject to error, bias, profiling, discrimination, and other unfair practices. This happens, in part, because algorithms are created by people, and people are not objective. Another reason this might happen is because there are historical biases and prejudices in the datasets AI systems learn from, and if people don't 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 should make when a crash is inevitable[4] racial, gender, and age biases in facial recognition; biases for people with disabilities; and AI making discriminatory decisions in the insurance and banking sectors. There are cases when people blame algorithms for injustices that negatively impact their lives, such as children being deprived of college admissions or being denied bail by judges reliant on automated systems.[5]

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 is it important for the purposes of this guide? Currently, numerous local government entities consider the use of facial recognition systems, without taking into account the risks and possible effects of their use, as this case illustrates.

See details in the Annex.

WHO IS RESPONSIBLE FOR AI OUTPUTS?

Humans create curate datasets, algorithms, consume the outputs of algorithms, and serve as role-models for them. Humans also provide an essential point of control for algorithms, either as testers or validators of the decisions made by algorithms. The responsibility for errors in AI decision making is a big moral concern. There are ongoing debates about who is responsible for any error or output of AI systems: the developer of the AI algorithm, the owner of the smart device, the operator of data, the person who provided input data or someone else? There are even debates if AI algorithms can be inventors or not. The lack of certainty on this front may continue increasing distrust, causing risks, and leading to long-lasting consequences for innovation and adoption of technology.

CASE-STUDY: IMPORTANCE OF DATA FOR FEMALE HEALTH

What happened: As AI becomes more widely used in the healthcare system, the lack of sufficient quality data about women's physiology and medical interventions and outcomes will reproduce current misdiagnoses.

Why is it important for the purposes of this guide? The common exclusion of women as test subjects in much medical research results in a lack of relevant data on women's health that leads to critical implications. This is often not considered by implementers.

See details in the Annex.

For reflection

How would machines solve the "Trolly Dilemma"?

4 "Trolly Dilemma" "A runaway trolley is barreling down the track and will kill five innocent people in the way. You can pull a switch that will direct the trolley to a different track. But another man is standing on that second track, and pulling the switch will lead to his death."

5 Jane Change (2021). Racism in, Racism out. A primer on algorithmic racism. Lilah Burke (2020). The Death and Life of an Admissions Algorithm. Pretrial Algorithms (Risk Assessment)

WHY DO SOME MORAL ISSUES APPEAR?

Applications of AI such as chatbots, criminal assessment systems, face recognition systems become sexist or racist 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 offers other examples of bias because these data sets may not include data on women, people with disabilities, indigenous peoples or other groups that are historically on the unfavorable side of the digital divide. In these cases, an AI algorithm would likely consider these populations as outliers and create models that lead to errors and ethical problems and those populations will continue to be invisible in the modern age.

For reflection

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

CASE-STUDY: BIASED RECRUITMENT AI SOLUTIONS

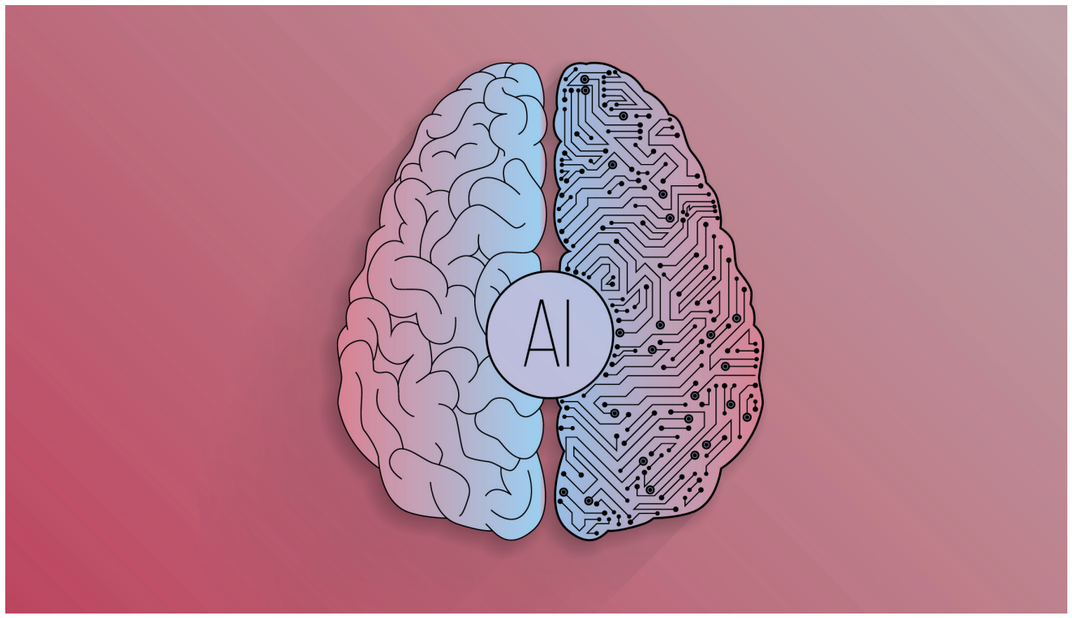
Country: USA

Year: 2018

What happened: Amazon abandoned its AI hiring program because of its bias against women because it trained on the resumes of the candidates who were mostly men

Why is it important for the purposes of this guide? Human Resources teams in both the private and public sectors are betting on the use of AI to streamline hiring processes, without taking into account case examples like this one and the negative effects of biased hiring.

See details in the Annex.



WHAT ARE THE RISKS?

Ethical problems can lead to a variety of consequences with different levels of severity. These include everything from increased inequality, to extended litigation processes to resist to social uprising. The profiling and biases of algorithms against a particular race or gender or specific category of people can affect how any system works, whether it is education or healthcare or finance, or even democracy. Core human values such as personal privacy, data protection, fairness, and autonomy become at risk. AI can be also used maliciously and cause damage in almost any field for faking data, stealing passwords, interfering with the work of other software and machines thus undermining trust in technology even more

WHO SHOULD BE INTERESTED IN AI ETHICS?

Everyone. Citizens, businesses, governments, academia are all users or adopters of AI applications and technologies. They all face ethical challenges that differently impact their lives. And they all have a perspective to add to the wider ethical AI conversation for our society.

CASE-STUDY: AI AGAINST PEOPLE WITH DISABILITIES

Country: Global

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

Why is it 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 to avoid that they may be excluded from the results or recommendations of such a system.

See details in the Annex.

ARE AI SYSTEMS BIASED GLOBALLY?

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. Most countries of the Global South still experience data poverty and are just forming reliable and robust digital infrastructures. Thus, there are concerns about whether western ideas of fairness should be considered universal[6] and if they can be applied in the same manner in developing countries as they are in advanced economies. Many question the primacy of western ethical traditions in most AI systems and wonder whether incorporation of ethical beliefs based on alternative systems inspired by, for instance Buddhism or Ubuntu, might change some assumptions about ethical AI. Yet, some nations do certain steps to change this trend. One example is Maori Principles for Data Sovereignty[7].

6 AI Decolonial Manifesto
7 Principles of Mori Data Sovereignty

HOW TO APPLY AI ETHICS

Governments around the globe have started to consider responses to both the opportunities and risks of AI. One method they have employed is the development of ethical guidelines for use of AI and other automated systems. While many of these guides are produced by institutions or governmental bodies without legal or legislative powers, the guides create a baseline of rules, begin to establish standards, and build awareness for government actors. Further, private sector organizations, academic institutions, and non-government organizations[8] have also created their own ethical principles or guidelines for a wider audience. Below you will find several options in terms of priority that can help foster the ethical use of AI:

1. CREATE AWARENESS AND INITIATE DIALOGUE

AI ethics are equally important for government, businesses, and individuals and should be discussed openly and widely. However, AI ethics is outside most peoples' awareness. Thus, a good starting point would be to launch awareness activities and bring all stakeholders to the table. It is important to ensure that there is the widest possible engagement and discussion on AI ethics issues, including national, sub-national, and municipal agencies, businesses (both large companies and small- and medium-sized enterprises), professional associations, and citizens.

CASE-STUDY: AI AGAINST PEOPLE WITH DISABILITIES

Country: Germany **Year:** 2018

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

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

See details in the Annex.

8 The booklet for start-ups and companies AI Ethics for Latin America by C Minds and Meta (2022).

2. ADOPT GUIDING PRINCIPLES

Many countries, cities, and organizations have developed principles and guidelines, codes of ethics, and self-assessment toolkits on AI for public officials and business organizations[9]. At the heart of each document there is a 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	Non-maleficence and trust	Equitability and fairness
AI uses open and understandable, auditable, and documented use of any data with clear up-to-date security measures	AI does no harm or inflict the least possible harm to reach an outcome. Continuous readiness to interfere in AI systems is ensured.	AI systems and solutions subordinate to human- defined rules and laws thus ensuring human rights above all

Non-maleficence and trust	Explainability
The AI system discloses who is responsible for AI works and solutions. At the same time, it refers that humans willingly give up some of decision-making power to AI	AI systems ensure the need to understand and hold to account the decision-making processes of AI

There are other principles adopted by different organizations and that are fixed in internationally recognized documents, for instance in 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 is it relevant for the purposes of this uide? Aspects and lessons learned from the Canadian experience can be taken into account 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 UN Moratorium on use of AI that threatens human rights or The fAIr LAC self-assessment tool (2022).

10 The Montréal Declaration for responsible AI development (2018).

11 The OECD AI Principles.

3. DEVELOP LEGAL FRAMEWORK FOR AI ETHICS

Almost all AI Ethics initiatives developed today have resulted in guidelines and frameworks that are recommended and 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: AI BILL OF RIGHTS

Country: USA **Year:** 2022

What happened: AI Bill of Rights was established and introduced protections individuals should have in the AI age.

Why is it relevant for the purposes of this guide? principles are presented that should guide the design, use and implementation of automated systems to protect citizens. These principles can be considered as a good practice to consider in the process of identifying applicable principles in local contexts.

See details in the Annex

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

AI ethics varies when it comes to specific sectors. Common principles of fairness and justice may not be enough when it comes to healthcare and medical diagnosis, e-commerce sales or education. It is pivotal to agree on a set of principles and guidelines relevant to different sectors. An agreement on this issue can be beneficial for the ubiquitous application of AI.

CASE-STUDY: LEARNING BOOK FOR UK AUTHORITIES

Country: UK **Year:** 2020

What happened: The government developed a Learning Book for Public Authorities to better understand how AI works and how it can be applied in the public sector.

Why is it relevant 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 axes in 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. The adoption of laws, norms and standards often does not keep pace with the introduction and use of technologies. Permanent and practical discussions around emerging ethical implications are important if one wants sustainable digital and data-driven transformation. Close collaboration with civil society could help monitor changes and ensure transparency.

CASE-STUDY: OPEN EXPLANATIONS OF AI FROM SINGAPORE

Country: Singapore **Year:** 2020

What happened: AI startup disclosed the exact parameters used in developing the 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 is it relevant 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.

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

Countries such as the UK, Canada, Germany, Japan, Argentina, United Arab Emirates, and international organizations such as the European Commission, OECD, GSMA and many others actively contribute to the AI Ethics agenda by issuing codes of ethics, guidelines, etc. or organizing consortiums such as the Global Partnership for AI (GPAI) to work together on values-based pathways and solutions for AI.[12] They are on the frontier of realizing how ethical adoption of AI can ensure inclusive and sustainable technological advancements and economic growth. Joining forces with international partners is important for facilitating trustworthy AI adoption.

CASE-STUDY: OECD AI GUIDELINES

Country: Global **Year:** 2019

What happened: OECD AI Principles – counts all 37 OECD countries and seven non-Member partners among its adherents, including Mexico, and has also formed the basis of G20 AI Principles.

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

See details in the Annex.

12 The Global Partnership of Artificial Intelligence (GPAI)

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), Aequitas bias and audit toolkit from Carnegie Mellon University, Fairlearn, Datasheets for Datasets from Microsoft, and Tensorflow Model Cards. 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 state-of-the-art algorithms to mitigate such bias and that is applied by many practitioners.[13] USAID also supports these endeavors. In 2021, it introduced Managing Machine Learning Projects in International Development, a practical toolkit for those in governments or public sectors that will work with data scientists to develop AI tools.[14]

13 Fairness 360
14 USAID Managing Machine Learning Projects in International Development: A Practical Guide

GLOSSARY

Definitions below are derived from multiple USAID studies[15], 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 - is 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.[16]

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.[17]

Machine Learning - is 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." This may include competencies that are variously referred to as computer literacy, information and communications technology 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.

15 USAID DECA Toolkit. USAID Digital Strategy. USAID Managing Machine Learning Projects in International Development: A Practical Guide

16 Chris Brook (2018). What is Cyber Hygiene? A Definition of Cyber Hygiene, Benefits, Best Practices, and More

17 UNESCO (2018). The Digital Literacy Global Framework (DLGF)

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. Many students have had their university places revoked as a result of the downgraded exam results. Almost 40% of the students received lower grades than they had anticipated and took to the streets and the courts for redress, forcing the government to retract the grades. Subsequent reviews suggested that the algorithms might have been biased (reinforcing prejudices in historical data, plus favoring smaller schools). Critics also took issue with the limited engagement and accountability tools that the government provided for students and parents. 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 the user interacts with things; 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 can also limit student participation and make them feel unsafe when they take responsibility for their ideas. Source.

FLAWED FACIAL RECOGNITION SYSTEMS

Google is one of the leaders in AI. But their facial recognition software is biased against African-Americans. In several stated cases, image recognition algorithms in Google Photos were classifying African-Americans as "gorillas". Simply preventing gorilla identification by image recognition algorithms is what the corporation has done, probably choosing to limit the service rather than run the risk of another 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 based around the 50th-percentile male. The most commonly used dummy is 1.77m tall and weighs 76kg (significantly taller and heavier than an average woman). This is one reason why women are 17 percent more likely to be killed and 47 percent more likely to be injured in crashes than men. One study of 12 of the most common fitness monitors found that they underestimated steps during housework by up to 74 percent (that was the Omron, which was within one percent for normal walking or running) and underestimated calories burned during housework by as much as 34 percent. Women are significantly excluded from medical research because the findings are not gender-disaggregated. The lack of disaggregated data means that half of the population is misrepresented or does not have equal access to public or private services or 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 of job postings 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 because 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. Besides, the controversy surrounding law enforcement's use of AI, significantly predictive policing, has led to sharp criticism because of its discriminatory effect. Source.

RACIST AI BOT IN TWITTER

In 2016 Microsoft launched an AI bot in Twitter that could interact and learn from users of the social media platform. However, it became racist and sexist a few hours after learning 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. Stereotypes sneak into translations because Google optimizes translations for English. As some experiments showed, in many cases, Google changes the gender of the word. For instance, the phrase "vier Historikerinnen und Historiker" (four male and female historians) is rendered as "cuatro historiadores" (four male historians) in Spanish, with similar results in Italian, French and Polish. Female historians are removed from the text. Source.

AI AND WOMEN'S HEALTH

The common exclusion of women as test subjects in much medical research results in a lack of relevant data on women's health. Heart disease, for example, has traditionally been thought of as a predominantly male disease, with "evidence-based" clinical guidelines based on male physiology. This has led to massive misdiagnosed or underdiagnosed heart disease in women. As AI becomes more widely used in the healthcare system, the lack of sufficient quality data about women's physiology and medical interventions and outcomes will reproduce current misdiagnoses. 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 that may mimic the direct and indirect discrimination they 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 northern province of Salta and the American tech giant Microsoft presented an algorithmic system to predict teenage pregnancy. They called it the Technology Platform for Social Intervention. The goal was to forecast which females from low-income areas would become pregnant over the following five years using the algorithm. The implications of being declared "predestined" for motherhood, or how knowing this would help prevent adolescent conception, were never made explicit. The system was based on 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, including 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 due to the complete absence of national AI legislation. No formal information about its accuracy or results has ever been released. Transparency and accountability are lacking. However, it was discovered that the system's database only contained information on racial and socioeconomic groups, not 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 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 a self-assessment tool. The latter helps government bodies assess the 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[18] was adopted in 2020. However, this initiative has not defined bias or mentioned gender bias as of this review. A new Blueprint for an AI Bill of Rights has introduced five protections individuals should have 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."[19] In October 2022, the Law on Artificial Intelligence Training for the Acquisition Workforce Act (also known as the AI Training Act)[20] was adopted to provide an AI training program for the acquisition workforce of executive agencies. The purpose of the program is to ensure the public workforce has knowledge of the capabilities and risks associated with AI. Several federal and state government agencies and the private sector have launched initiatives to prevent algorithmic bias in the public and private sectors. 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 start-up 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, UCARE.AI 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 Parkway's data managers and its patients. Source.

18 National Artificial Intelligence Initiative Act of 2020.

19 Algorithmic Discrimination Protections (2020)

20 Artificial Intelligence Training for the Acquisition Workforce Act or the AI Training Act (2022).

LEARNING BOOK FOR PUBLIC AUTHORITIES IN THE UK

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

AI ETHICS – A JOINT WORK

Recognizing that issues relevant to AI transcend borders, countries are also 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.

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