What is "ethical AI"? Leading or participating on an ethical team and/or working in statistics, data science, and artificial intelligence

Rochelle E. Tractenberg

Collaborative for Research on Outcomes and –Metrics; and Departments of Neurology; Biostatistics, Bioinformatics & Biomathematics; and Rehabilitation Medicine, Georgetown University, Washington, DC, USA

ORCID: 0000-0002-1121-2119

Correspondence to:
Rochelle E. Tractenberg
Georgetown University Neurology
Room 207, Building D
4000 Reservoir Rd., NW
Washington, DC, 20057 USA
Email: rochelle -dot- tractenberg -at -gmail -dot- com

Acknowledgement: There are no actual or potential conflicts of interest. Opinions expressed in this article are the author's own.

Running Head: What is Ethical AI?

To appear in: In, H. Doosti, (Ed.). *Ethical Statistics.* Cambridge, UK: Ethics International Press.
Preprint available at: *StatArXiv* TBD

Abstract
Artificial Intelligence (AI) arises from computing and statistics, and as such, can be developed and deployed ethically when the ethical practice standards of each of these fields are followed. *The Toronto* Declaration was formulated in 2018 specifically to ensure that machine learning and AI could be held accountable for respecting, and promoting, universal human rights. The *Code of Ethics and Professional Conduct* of the Association of Computing Machinery (ACM, 2018) and the *Ethical Guidelines for Statistical Practice* of the American Statistical Association (ASA, 2022) describe the ethical practice standards for any person at any level of training or job title who utilizes computing (ACM) or statistical practices (ASA). These three reference documents can together define "what is ethical AI". All development, deployment, and use of computing is covered by the ACM *Code*; the ASA defines statistical practice to "include activities such as: designing the collection of, summarizing, processing, analyzing, interpreting, or presenting, data; as well as model or algorithm development and deployment." Just as the *Toronto Declaration* describes universal human rights protections, the ACM and ASA ethical practice standards apply to professionals, individuals with diverse background or jobs that include computing and statistical practices at any point, and employers, clients, organizations, and institutions that employ or utilize the outputs from computing and statistical practices worldwide. The ACM *Code of Ethics* has four Principles, including one specifically for Leaders with seven elements. The ASA *Ethical Guidelines* include eight principles and an Appendix; one Guideline Principle (G. Responsibilities of Leaders, Supervisors, and Mentors in Statistical Practice) with its five elements and the Appendix (Responsibilities of organizations/institutions) with its 12 elements are specifically intended to support workplace engagement with, and support of, ethical statistical practices, plus, the specific roles and responsibilities of those in leadership positions. These ethical practice standards can support both individual practitioners', and leaders', meeting their obligations for ethical AI worldwide.

Keywords: ethical leadership; ethical statistical practice; ethical computing; ethical AI; international ethical guidelines for statistical practice; ethical artificial intelligence; data ethics.

1. Introduction

Artificial Intelligence (AI) is an application arising from the disciplines of computing and statistics, and as such, can be developed and deployed ethically when the ethical practice standards of each of these fields are followed. *The Toronto* Declaration was formulated in 2018 specifically to ensure that machine learning and AI could be held accountable for respecting, and promoting, universal human rights. The *Code of Ethics and Professional Conduct* of the Association of Computing Machinery (ACM, 2018) The *Ethical Guidelines for Statistical Practice* of the American Statistical Association (ASA, 2022) describe the ethical practice standards for any person at any level of training or job title who utilizes computing (ACM) or statistical practices (ASA). These three reference documents can together define "what is ethical AI".

Cross- and multi-disciplinary teams working with computational practitioners, statisticians, data scientists, and developers and deployers of AI are increasingly common. Leaders of such teams need to understand that creating a culture of ethical practice can be done in a straightforward, reproducible, way. From the leaders' perspective, establishing a culture that supports ethical practice includes promoting thoughtful decisions about data as well as practice. A commitment to ethical practice will be important to both recruiting and retaining talent. Leaders have both unique responsibilities and a unique opportunity to encourage, promote, and engage in ethical statistical and data science practice.

1.1 What is AI

"…artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving. It also encompasses sub-fields of machine learning and deep learning... AI algorithms … seek to create expert systems which make predictions or classifications based on input data."[1]

Related terms are "machine learning" and "statistical learning". According to Wikipedia, "Machine learning is a branch of statistics and computer science". Similarly, "statistical learning theory is a framework for machine learning drawing from the fields of statistics and functional analysis. (it)… deals with the statistical inference problem of finding a predictive function based on data…The goals of (statistical) learning are understanding and prediction."

Therefore, AI, whether or not it is defined to include data science, machine learning, or statistical learning, arises at the intersection of computing and statistics. There may be some types of AI that do not involve statistical practices, defined in the next section, but this chapter assumes that the majority of AI applications will involve statistical practices.

[1] https://www.ibm.com/topics/artificial-intelligence#:~:text=At%20its%20simplest%20form%2C%20artificial,in%20conjunction%20with%20artificial%20intelligence.

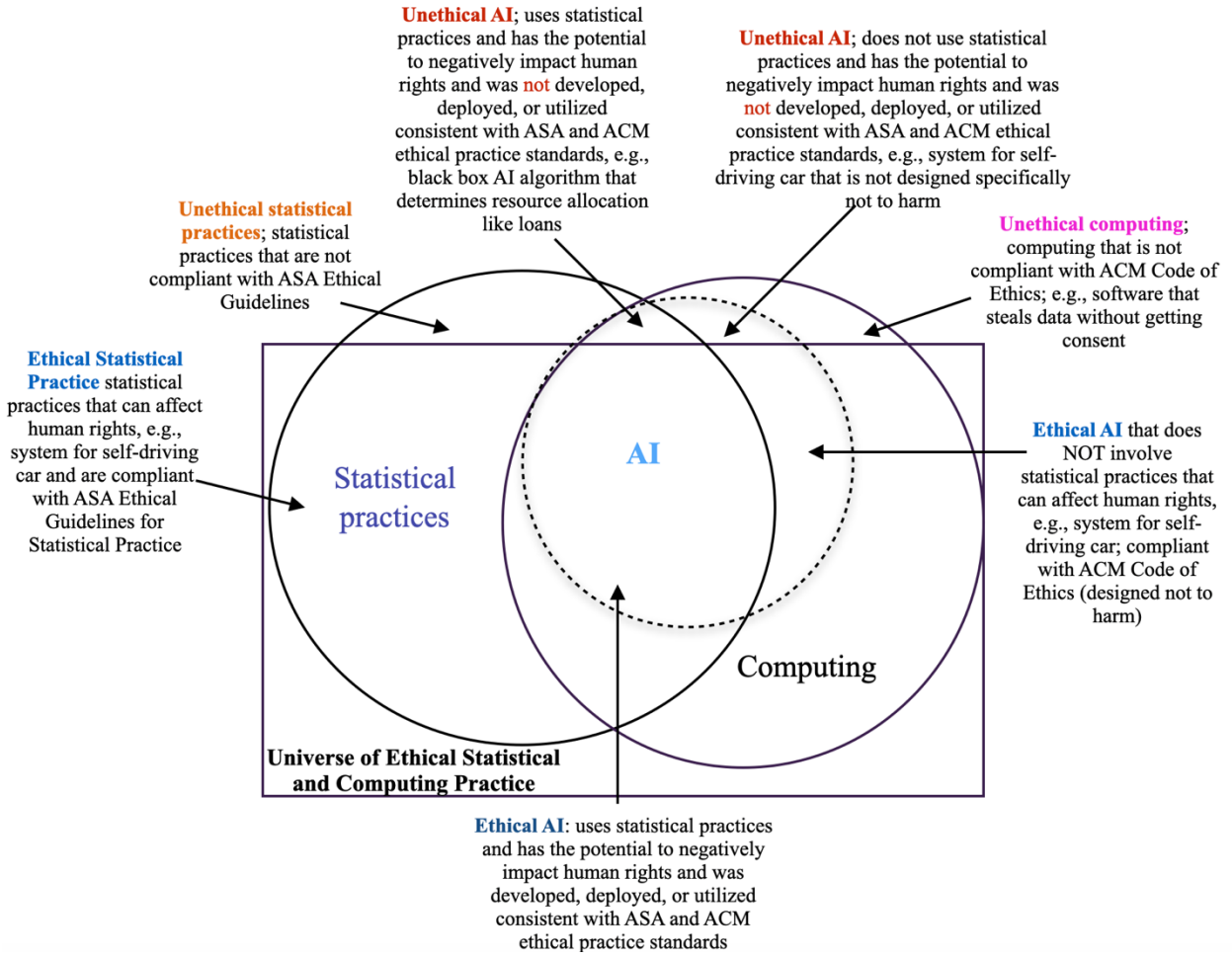
With that in mind, the following sections discuss how "ethical AI" emerges from ethical computing and ethical statistics and data science.

1.2 What does "ethical AI" mean?

There are three dimensions of "ethical AI". The first dimension reflects fundamental human rights: the *Toronto Declaration* (2018) outlines the requirements for "ethical AI" by state and private sector actors. Specifically, "States have obligations to **promote, protect and respect** human rights; private sector actors, including companies, have a responsibility to **respect** human rights at all times." (emphasis added). Thus, AI applications that subvert or undermine these obligations, and those that do not promote, protect, and respect (states) or respect (private sector) fundamental human rights, cannot be characterized as "ethical". This is true whether the AI itself undermines the responsibilities, or the uses to which the AI or its outputs are put undermine these responsibilities.

The second dimension of "ethical AI" is that, if the AI uses computing in an impactful way, or the AI itself is computing to be used in an impactful way, then actual or potential harms of that AI must be recognized and minimized. This is articulated by the Association of Computing Machinery *Code of Ethics and Professional Conduct* (2018). The third dimension of "ethical AI" is the reliance on statistical practices. The American Statistical Association (ASA, 2022) specifies, "“statistical practice” includes activities such as: designing the collection of, summarizing, processing, analyzing, interpreting, or presenting, data; as well as model or algorithm development and deployment. Throughout these Guidelines, the term "statistical practitioner" includes all those who engage in statistical practice, regardless of job title, profession, level, or field of degree. The Guidelines are intended for individuals, but these principles are also relevant to organizations that engage in statistical practice." Given this definition of statistical practices, many if not most AI applications will rely on statistical practices. The three dimensional definition of ethical AI is reflected in Figure 1.

Figure 1. How ethical AI arises from ethical statistical and ethical computing practices.



As suggested by Toronto, and echoed by many scholars (e.g., Latonero, 2018), it is possible to define "ethical AI" in a concretely achievable way. Throughout the design, development, deployment, and utlilization of any AI product, it is possible, as well as urgently needed, to incorporate what Latonero characterizes as "Human Rights Impact Assessments" (p. 2). Table 1 outlines the responsibilities for assuring universal human rights, which pertain generally but are specifically relevant in considerations of "ethical AI".

Table 1. The *Toronto Declaration* (2018) outline of universal human rights and obligations towards these for State and Private Sector Actors.

	Toronto Declaration components: "States have obligations to **promote, protect and respect** human rights; private sector actors, including companies, have a responsibility to **respect** human rights at all times." (emphasis added)

ACTOR:	The right to equality and non-discrimination	Preventing discrimination	Protecting the rights of all individuals and groups: promoting diversity and inclusion	Human rights due diligence: i. Identify potential discriminatory outcomes ii. Take effective action to prevent and mitigate discrimination and track responses iii. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems.	Accountability of individuals and organizations (including business entities, governments, and multi-country entities)
STATE:	Promote and protect and respect the rights to equality and non-discrimination in every state-created or state-deployed instance.	Promote nondiscrimination, protect against discrimination, and respect individuals' right to non-discrimination.	Promote diversity & inclusion, protect against exclusion, and respect individuals' right to be included.	Promote human rights due dilligence, protect against actors who fail to competently and reliably exercise human rights due dilligence, and respect individuals' right to human rights due dilligence in every state-created or state-deployed instance.	Promote, and accept accountability, for respecting these human rights, and for exercising human rights due dilligence, by all actors (state and non-state).

PRIVATE SECTOR:	Respect individuals' rights to equality and non-discrimination AND respect state efforts and responsibilities to promote respect, and protect against violations of these rights, AND respect other (non-state) groups' efforts to promote/protect these rights.	Respect state efforts and responsibilities to promote nondiscrimination, and protect against discrimination, AND respect other (non-state) groups' efforts to prevent discrimination.	Respect state efforts and responsibilities to promote diversity & inclusion, protect against exclusion, and respect individuals' right to be included AND respect other (non-state) groups' efforts to promote diversity & inclusion, and to protect against exclusion.	Universally and continuously exercise human rights due diligence in every instance in an ongoing manner.	Promote and accept accountability for respecting these human rights, and for exercising human rights due dilligence, by all actors (state and non-state).

The full *Toronto Declaration* appears in the Appendix. The *Declaration* considers only those harms to human rights to equality, non-discrimination, diversity, and inclusion. It defines its scope: "Discrimination is defined under international law as “any distinction, exclusion, restriction or preference which is based on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, and which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise by all persons, on an equal footing, of all rights and freedoms” (United Nations, 1989). This list is non-exhaustive as the United Nations High Commissioner for Human Rights has recognized the necessity of preventing discrimination against additional classes" (United Nations, 2017). Echoing and elaborating the universal rights considered in the Toronto Declaration, the Council of Europe (2020) specifies "significant human rights challenges attached to the increasing reliance on algorithmic systems in everyday life, such as regarding the right to a fair trial; the right to privacy and data protection; the right to freedom of thought, conscience and religion; the right to freedom of expression; the right to freedom of assembly; the right to equal treatment; and economic and social rights."

Importantly, the Institute of Electrical and Electronics Engineers (IEEE) initiated a global consensus project on "(t)he ethical design, development, and implementation of ...intelligent and autonomous technical systems" (IEEE 2017, p.6). This consensus document, the result of the Global Initiative on Ethics of Autonomous and Intelligent Systems, is called "Ethically Aligned Design". The second version (2017) represents the consensus of key stakeholder representatives, "several hundred participants from six continents, who are thought leaders from academia, industry, civil society, policy and government " (p. 3) While the *Toronto Declaration* sought to outline the legal obligations

of state and private sector actors in terms of universal human rights, the IEEE project's is to "establish *frameworks to guide and inform dialogue and debate* around the non-technical implications of these technologies." (emphasis added; p.6). Version 2 of *Ethically Aligned Design* (EAD) preceded the *Toronto Declaration*, and the general principles of EAD are tightly coordinated with *Toronto*, specifying that "...ethical design, development, and implementation of intelligent and autonomous technical systems should be guided by:
Human Rights: Ensure they do not infringe on internationally recognized human rights
Well-being: Prioritize metrics of well-being in their design and use
Accountability: Ensure that their designers and operators are responsible and accountable
Transparency: Ensure they operate in a transparent manner
Awareness of misuse: Minimize the risks of their misuse"

While EAD and the *Toronto Declaration* are strongly aligned in the articulation of obligations by designers, developers, and implementers of intelligent and autonomous systems and AI to protect and respect human rights, they do not offer specific guidelines - nor were they intended to do so (see, e.g., Novelli et al. 2023). The ethical practitioner in AI - which includes leaders of teams working on, in, and with AI, needs more concrete and actionable guidance. Fortunately, both computing and statistics and data science, the two core components of AI, offer exactly that guidance. Moreover, the "significant human rights" alluded to in the Council of Europe declaration includes specific rights that are protected by ethical computing and ethical statistical practices; thus, in addition to Human Rights Impact Assessments (Latonero, 2018), ethical practitioners can also assess their compliance with the specifically relevant ethical practice standards for computing (ACM, 2018) and statistics and data science (ASA, 2022). These are the focus of the next two sections.

Ethics is defined as "the principles of conduct governing an individual or a group (e.g., professional ethics)"[2]. Because individuals or groups who engage in developing, deploying, or utilizing the outputs of AI are utilizing statistical and computing practices, the ethical practice standards that derive from statistics and data science(American Statistical Association (ASA), 2022), computing (Association of Computing Machinery (ACM), 2018) pertain to any and all AI activities. Gillikin (2017) defines a “practice standard” as a document to “define the way the profession’s body of knowledge is ethically translated into day-to-day activities” (Gillikin 2017, p. 1). Critically, both the ASA and ACM assert the applicability of their ethical practice standards for *all* who utilize their domain knowledge, skills, and technologies. Although there may be many and varied "codes" or oaths relating to "ethical AI" (e.g., Jobin et al. 2019; Mittelstadt, 2019), because they are neither comprehensive, consensus-based, nor actually very different from the professional ethical practice standards. Thus, *whenever an individual - irrespective of membership in these professional organizations, degree or training, or job title- uses statistical practices or computing, the ASA and ACM ethical practice*

[2] https://www.merriam-webster.com/dictionary/ethics

standards are relevant. Therefore, considerations of "ethical AI" must include these ethical practice standards; they appear in full in the Appendix. Alignment of ethical practices in computing and statistics and data science with the fundamental human rights outlined in the Toronto Declaration, outlined in the next sections, is achieved following the Degrees of Freedom Analysis Method (Tractenberg, 2023).

1.3 Association of Computing Machinery (ACM) *Code of Ethics and Professional Conduct* (2018)

in the Preamble, the ACM Code of Ethics asserts: "*Computing professionals' actions change the world. To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good. ... The Code is designed to inspire and guide the ethical conduct of all computing professionals, including current and aspiring practitioners, instructors, students, influencers, and anyone who uses computing technology in an impactful way*."

Under Principle 1 (*General Moral Principles*), ACM Code element **1.1** states:

> This principle, which concerns the quality of life of all people, affirms an obligation of computing professionals, both individually and collectively, to use their skills for the benefit of society, its members, and the environment surrounding them. This obligation includes promoting fundamental human rights and protecting each individual's right to autonomy. An essential aim of computing professionals is to minimize negative consequences of computing, including threats to health, safety, personal security, and privacy. When the interests of multiple groups conflict, the needs of those less advantaged should be given increased attention and priority.
>
> Computing professionals should consider whether the results of their efforts will respect diversity, will be used in socially responsible ways, will meet social needs, and will be broadly accessible. They are encouraged to actively contribute to society by engaging in pro bono or volunteer work that benefits the public good.
>
> In addition to a safe social environment, human well-being requires a safe natural environment. Therefore, computing professionals should promote environmental sustainability both locally and globally.

The second element in Principle 1 of the Code is similarly applicable to ethical AI:

> **1.2**: Avoiding harm begins with careful consideration of potential impacts on all those affected by decisions. When harm is an intentional part of the system, those responsible are obligated to ensure that the harm is ethically justified. In either case, ensure that all harm is minimized... A computing professional has an

> additional obligation to report any signs of system risks that might result in harm. If leaders do not act to curtail or mitigate such risks, it may be necessary to “blow the whistle” to reduce potential harm.

As noted, the qualitative alignment appearing in Tables 2 and 3 was carried out following the Degrees of Freedom Analysis method outlined in Tractenberg (2023). Where there is direct alignment between, or support for, a *Toronto Declaration* responsibility with an ACM element, an "x" appears in the cell. If the alignment is indirect or implicit, the cell contains "(x)". If the language is not explicitly or implicitly aligned, the cell is blank. Table 2 identifies the specific elements of the ACM *Code of Ethics* that enable the ethical computing professional (or user of computing) to meet or exceed their obligations under the *Toronto Declaration*. The Full *ACM Code of Ethics and Professional Conduct* (2018) appears in the Appendix.

Table 2. Elements of the Association of Computing Machinery *Code of Ethics* (2018) that facilitate meeting responsibilities under the *Toronto Declaration* (2018).

	Toronto Declaration components: "States have obligations to **promote, protect and respect** human rights; private sector actors, including companies, have a responsibility to **respect** human rights at all times." (emphasis added)				
Toronto Declaration: **ACM:**	The right to equality and non-discrimination	Preventing discrimination	Protecting the rights of all individuals and groups: promoting diversity and inclusion	Human rights due diligence: i. Identify potential discriminatory outcomes ii. Take effective action to prevent and mitigate discrimination and track responses iii. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems.	Accountability of individuals and organizations (including business entities, governments, and multi-country entities)
ACM 1. GENERAL MORAL PRINCIPLES. ***A computing professional should…***					
1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.	x		(x)	(x)	(x)
1.2 Avoid harm. *In this document, "harm" means negative consequences to any stakeholder, especially when those consequences are significant and unjust.*	x	x	x	x	x

1.3 Be honest and trustworthy.					(x)
1.4 Be fair and take action not to discriminate.	x	x	(x)	x	x
1.5 Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.					
1.6 Respect privacy.					
1.7 Honor confidentiality.					
2. PROFESSIONAL RESPONSIBILITIES. ***A computing professional should…***					
2.1 Strive to achieve high quality in both the process and products of professional work.					
2.2 Maintain high standards of professional competence, conduct, and ethical practice.	x	x		x	x
2.3 Know, respect, and apply existing rules pertaining to professional work.	x	x		x	x
2.4 Accept and provide appropriate professional review.					
2.5 Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks.		x		x	x

2.6 Have the necessary expertise, or the ability to obtain that expertise, for completing a work assignment before accepting it. Once accepted, that commitment should be honored.					
2.7 Improve public awareness and understanding of computing, related technologies, and their consequences.					
2.8 Access computing and communication resources only when authorized to do so.					(x)
2.9 Design and implement systems that are robustly and usably secure.					
3. PROFESSIONAL LEADERSHIP PRINCIPLES. *In this section, "leader" means any member of an organization or group who has influence, educational responsibilities, or managerial responsibilities. These principles generally apply to organizations and groups, as well as their leaders.* ***A computing professional should…***					
3.1 Ensure that the public good is the central concern during all professional computing work.	x	x	x	x	(x)
3.2 Articulate, encourage acceptance of, and evaluate fulfillment of the social responsibilities of members of an organization or group.					

3.3 Manage personnel and resources to enhance the quality of working life.					
3.4 Articulate, apply, and support policies and processes that reflect the principles in the Code.				x	x
3.5 Create opportunities for members of the organization or group to learn and be accountable for the scope, functions, limitations, and impacts of systems.					
3.6 Retire legacy systems with care.					
3.7 Recognize when a computer system is becoming integrated into the infrastructure of society, and adopt an appropriate standard of care for that system and its users.	(x)	(x)	(x)	x	x
4. COMPLIANCE WITH THE CODE. ***A computing professional should...***					
4.1 Uphold, promote, and respect the principles of the Code.				(x)	(x)
4.2 Treat violations of the Code as inconsistent with membership in the ACM.				(x)	(x)

Table 2 outlines the varieties of responsibilities that individual computing professionals, as well as those in leadership roles, should intentionally follow. Note that privacy - defined as "the state of being free from unwarranted intrusion into the private life of individuals" (Office of Management and Budget, 2020, p. 13)- and confidentiality - defined as "the state of one's information being free from inappropriate access and use" (Office of Management and Budget, 2020, p. 13)- are both specifically protected in ACM elements 1.6 and 1.7, and these are not specifically identified in the *Toronto Declaration*. An individual who complies with the ethical computing practice standard will still respect these particular rights, and, there is no rationale for assertions that an expectation or legal requirement that AI and those who design, develop, deploy, or utilize the outputs of AI who respect these rights encounter conflict with rights to free speech (e.g., Balkin, 2016). While Table 1 shows the general responsibilities of ethical computing professionals, the following two elements of the ACM *Code of Ethics* are particularly important for those in leadership roles (emphasis added):

> **1.1**: This principle, which concerns the quality of life of all people, affirms an obligation of computing professionals, both individually and collectively, to use their skills for the benefit of society, its members, and the environment surrounding them. **This obligation includes promoting fundamental human rights and protecting each individual's right to autonomy. An essential aim of computing professionals is to minimize negative consequences of computing**, including threats to health, safety, personal security, and privacy. When the interests of multiple groups conflict, the needs of those less advantaged should be given increased attention and priority.
>
> **Computing professionals should consider whether the results of their efforts will respect diversity, will be used in socially responsible ways, will meet social needs, and will be broadly accessible**. ...
>
> In addition to a safe social environment, human well-being requires a safe natural environment. Therefore, **computing professionals should promote environmental sustainability both locally and globally**. (emphasis added)
>
> **1.2:** Avoiding harm begins with careful consideration of potential impacts on all those affected by decisions. When harm is an intentional part of the system, those responsible are obligated to ensure that the harm is ethically justified. In either case, ensure that all harm is minimized... A computing professional has an additional obligation to report any signs of system risks that might result in harm. If leaders do not act to curtail or mitigate such risks, it may be necessary to "blow the whistle" to reduce potential harm.
>
> **1.4:** Be fair and take action not to discriminate.

Each of these elements of *the Code* pertain to individuals as well as leaders; leaders have an additional (implicit) obligation to ensure that practitioners in their groups and teams, as well as the end users and stakeholders who are affected by computing practitioner decisions, follow these *Code* elements in order to be considered to be practicing, developing, and deploying, "ethical AI".

1.4 ASA Ethical Guidelines for Statistical Practice (2022):

The Preamble to the *Ethical Guidelines* states that:

> ""statistical practice" includes activities such as: designing the collection of, summarizing, processing, analyzing, interpreting, or presenting, data; as well as model or algorithm development and deployment. Throughout these Guidelines, the term "statistical practitioner" includes all those who engage in statistical practice, regardless of job title, profession, level, or field of degree. The Guidelines are intended for individuals, but these principles are also relevant to organizations that engage in statistical practice."
>
> The Ethical Guidelines aim to promote accountability by informing those who rely on any aspects of statistical practice of the standards that they should expect."

Principle A, **Professional Integrity and Accountability**: "Ethical statistical practice supports valid and prudent decision making with appropriate methodology."

This fundamental Principle of ethical statistical and data science practice is of critical importance in AI applications because of the focus on decision making. In addition to ensuring that the fundamental human rights outlined in the *Toronto Declaration* are respected, Principle A explicitly charges the ethical practitioner with accountability for the decisions that are rendered by the statistical practices embedded in AI applications.

Element A2 is even more specific in its charges for the ethical practitioner: **A2**: "(The ethical statistical practitioner) Uses methodology and data that are valid, relevant, and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results." When an AI application renders a decision - e.g., about a loan, sentencing, or recommendations for purchase, the developer and user of that AI application must consider whether the decision is "valid, interpretable, and reproducible". If the same algorithm is used, with the same inputs (varying only according to the individual), to render a decision then the AI application is valid, interpretable, and reproducible. If an AI application uses an algorithm that cannot reproduce any given decision, then it cannot be determined to be using "methodology and data that are valid, relevant, and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results." An auditable, but irreproducible, AI application wil not meet the test of ethical AI. If the AI application decision is about purchase recommendations, the potential risk of violating fundamental and universal human rights might be minimal. However, when an AI

application renders a decision that can adversely affect individuals, the decisions must be reproducible or else the application cannot be considered "ethical AI".

Ethical AI requires that those who develop, deploy, or utilize AI must "support valid and prudent decision making with appropriate methodology", without violating or risking violation of the responsibilities for universal human rights outlined in the *Toronto Declaration.*

Further, ASA *Ethical Guideline* Principle B, **Integrity of Data and Methods** states that the ethical practitioner "communicates potential impacts (of data and methods) on the interpretation, conclusions, recommendations, decisions, or other results of statistical practices." To the extent that an AI application generates or is intended to support interpretation, conclusions, recommendations, decisions, or other results that can violate or risk violation of the responsibilities for universal human rights outlined in the *Toronto Declaration,* ASA Principle B describes distinct responsibilities that must be met, or else the application cannot be considered "ethical AI".

Like the ACM, the ASA is also concerned with stakeholders in the decisions that are made, based on, or supported by practitioners in the field. ASA Principle C, **Responsibilities to Stakeholders,** states: "Those who fund, contribute to, use, or are affected by statistical practices are considered stakeholders. The ethical statistical practitioner respects the interests of stakeholders while practicing in compliance with these *Guidelines*." Ethical AI requires that the responsibilities to various stakeholders are prioritized such that there is no, or only minimal, risk of violating fundamental and universal human rights.

The *Toronto* Declaration is technically focused on machine learning, which is specifically informed by data, which is the topic of ASA *Ethical Guidelines* Principle D, **Responsibilities to Research Subjects, Data Subjects, or those directly affected by statistical practices**. Specifically, Principle D states, "The ethical statistical practitioner does not misuse or condone the misuse of data. They protect and respect the rights and interests of human and animal subjects. These responsibilities extend to those who will be directly affected by statistical practices." Element D11 articulates clearly how crucial it is to follow the ASA Ethical Guidelines in order to engage in ethical AI: "(The ethical statistical practitioner) Does not conduct statistical practice that could reasonably be interpreted by subjects as sanctioning a violation of their rights. Seeks to use statistical practices to promote the just and impartial treatment of all individuals."

These are examples of the specific guidance the ASA *Ethical Guidelines for Statistical Practice* afford to those who develop, deploy, or utilize the outputs of AI to ensure that AI - and any statistical practices - do not directly or indirectly violate, or pose a risk to, fundamental human rights. Additional elements from the ASA *Ethical Guideline*s that are relevant for ensuring ethical AI, and any AI or machine learning that will not violate universal human rights, appear in Table 3.

Table 3. Elements of the *ASA Ethical Guidelines for Statistical Practice* (2022) that facilitate meeting responsibilities under the *Toronto Declaration* (2018).

	Toronto Declaration components: "States have obligations to **promote, protect and respect** human rights; private sector actors, including companies, have a responsibility to **respect** human rights at all times." (emphasis added)				
ASA Ethical Guideline Principles	The right to equality and non-discrimination	Preventing discrimination	Protecting the rights of all individuals and groups: promoting diversity and inclusion	Human rights due diligence: i. Identify potential discriminatory outcomes ii. Take effective action to prevent and mitigate discrimination and track responses iii. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems.	Accountability of individuals and organizations (including business entities, governments, and multi-country entities)

A. Professional Integrity & Accountability: Professional integrity and accountability require taking responsibility for one's work. Ethical statistical practice supports valid and prudent decision making with appropriate methodology. The ethical statistical practitioner represents their capabilities and activities honestly, and treats others with respect. (12 elements)	A3	A3 (A4) A8			A11
B. Integrity of data and methods: The ethical statistical practitioner seeks to understand and mitigate known or suspected limitations, defects, or biases in the data or methods and communicates potential impacts on the interpretation, conclusions, recommendations, decisions, or other results of statistical practices. (7 elements)				B1, B2, B3, B4	B1 (B3) B4
C. Responsibilities to Stakeholders: Those who fund, contribute to, use, or are affected by statistical practices are considered		(C2)	(C2)	C2	C2

stakeholders. The ethical statistical practitioner respects the interests of stakeholders while practicing in compliance with these Guidelines. (8 elements)					
D. Responsibilities to research subjects, data subjects, or those directly affected by statistical practices: The ethical statistical practitioner does not misuse or condone the misuse of data. They protect and respect the rights and interests of human and animal subjects. These responsibilities extend to those who will be directly affected by statistical practices. (11 elements)	D1, D6, D11	D1, D6	D1, D11	D1, D6, D8, D10, D11	D1, (D5), D8, D10, D11
E. Responsibilities to members of multidisciplinary teams: Statistical practice is often conducted in teams made up of professionals with different professional standards. The statistical practitioner must know how to work ethically in this environment. (4 elements)		(E4)		E3, E4	E3, E4

F. Responsibilities to Fellow Statistical Practitioners and the Profession: Statistical practices occur in a wide range of contexts. Irrespective of job title and training, those who practice statistics have a responsibility to treat statistical practitioners, and the profession, with respect. Responsibilities to other practitioners and the profession include honest communication and engagement that can strengthen the work of others and the profession. (5 elements)	F4	F4	F4	F4	F4
G. Responsibilities of Leaders, Supervisors, and Mentors in Statistical Practice: Statistical practitioners leading, supervising, and/or mentoring people in statistical practice have specific obligations to follow and promote these Ethical Guidelines. Their support for – and insistence on – ethical statistical practice are essential for the integrity of the practice and	G1 (G5)	G1 (G5)	G1 (G5)	G1 (G5)	G1 (G5)

profession of statistics as well as the practitioners themselves. (5 elements)					
H. Responsibilities regarding potential misconduct: The ethical statistical practitioner understands that questions may arise concerning potential misconduct related to statistical, scientific, or professional practice. At times, a practitioner may accuse someone of misconduct, or be accused by others. At other times, a practitioner may be involved in the investigation of others' behavior. Allegations of misconduct may arise within different institutions with different standards and potentially different outcomes. The elements that follow relate specifically to allegations of statistical, scientific, and professional misconduct. (8 elements)				H2	H2
APPENDIX: Responsibilities of organizations/institutions: Whenever organizations and institutions design the collection of, summarize,	App4, App 6				App8, App10

process, analyze, interpret, or present, data; or develop and/or deploy models or algorithms, they have responsibilities to use statistical practice in ways that are consistent with these Guidelines, as well as promote ethical statistical practice. (Organizations 7 elements; Leaders 5 elements; 12 elements total)					

Table 3 highlights the elements of each Principle, and the Appendix, of the *ASA Ethical Guidelines for Statistical Practice* that is relevant for individuals who practice, or utilize the outputs from, statistical practices. The reader who explores the *Ethical Guidelines for Statistical Practice* will see specific responsibilities of statistics practitioners to protect and respect confidentiality (B4; D4, D7, D9) and privacy (D4, D9). Thus, like with the ACM *Code of Ethics*, ethical statistical practitioners who follow the ASA *Ethical Guidelines* will respect and promote these particular rights, even though they are not specifically included in the *Toronto Declaration*.

Also shown in Table 3 is specific guidance from Principle G and the Appendix that are important to consider on their own. Principle G and the Appendix elements appear below.

Principle G (Responsibilities of Leaders, Supervisors, and Mentors in Statistical Practice) specifies, "Statistical practitioners leading, supervising, and/or mentoring people in statistical practice have specific obligations to follow and promote these Ethical Guidelines. Their support for – and insistence on – ethical statistical practice are essential for the integrity of the practice and profession of statistics as well as the practitioners themselves." The two elements of Principle G that are most strongly supportive of meeting the obligations outlined in the *Toronto Declaration* are:

> **Those leading, supervising, or mentoring statistical practitioners are expected to**:
> G1: Ensure appropriate statistical practice that is consistent with these Guidelines. Protect the statistical practitioners who comply with these Guidelines, and advocate for a culture that supports ethical statistical practice.
> G5: Establish a culture that values validation of assumptions, and assessment of model/algorithm performance over time and across relevant subgroups, as needed. Communicate with relevant stakeholders regarding model or algorithm maintenance, failure, or actual or proposed modifications.

The **Appendix** to the ASA *Ethical Guidelines for Statistical Practice* (Responsibilities of organizations/institutions) states, "Whenever organizations and institutions design the collection of, summarize, process, analyze, interpret, or present, data; or develop and/or deploy models or algorithms, they have responsibilities to use statistical practice in ways that are consistent with these Guidelines, as well as promote ethical statistical practice." The Appendix has two components, with two elements each that are highly relevant to leading ethical teams in AI, and, to ensuring that AI that is developed, deployed, or used to make decisions by an organization or institution is done so in an ethical manner:

Organizations and institutions engage in, and promote, ethical statistical practice by:

App 4: Supporting statistical practice that is objective and transparent. Not allowing organizational objectives or expectations to encourage unethical statistical practice by its employees.
App 6: Avoiding statistical practices that exploit vulnerable populations or create or perpetuate discrimination or unjust outcomes. Considering both scientific validity and impact on societal and human well-being that results from the organization's statistical practice.

Those in leadership, supervisory, or managerial positions who oversee statistical practitioners promote ethical statistical practice by following Principle G and:

App 8: Recognizing that it is contrary to these Guidelines to report or follow only those results that conform to expectations without explicitly acknowledging competing findings and the basis for choices regarding which results to report, use, and/or cite.
App 10: In cases where ethical issues are raised, representing them fairly within the organization's leadership team.

2. Discussion

The ACM *Code of Ethics* and the ASA *Ethical Guidelines for Statistical Practice* are critical facilitators of ethical AI, as shown in Figure 1. Moreover, they contain specific support for individual practitioners and leaders alike to meet responsibilities to protect and promote universal human rights as outlined in the *Toronto Declaration*. It is important to note that the *Toronto Declaration* was drafted to specifically represent the legal underpinnings for articulating accountability for state and private sector actors in terms of how AI affects humans worldwide. These fundamental universal human rights do not include abstract constructs like "peace" and "beneficence" (e.g., Ryan & Stahl, 2020) -which are difficult if not impossible to measure. The ethical practice standards for computing (ACM) and statistical practices (ASA) are each intended to promote ethical decision making by practitioners and users of outputs of statistics, data science, and computing, irrespective of job title, level, or field of degree. Like the obligation to respect human rights, the responsibilities to follow ethical practice standards pertain whenever individuals - or their programs or AI - engage in or utilize statistical practice, which includes designing the collection of, summarizing, processing, analyzing, interpreting, or presenting, data; as well as model or algorithm development and deployment. To the extent that AI is not developed, deployed, or utilized following the ACM and ASA ethical practice standards, it cannot be considered or characterized to be "ethical AI". Designing, developing, deploying, implementing, and utilizing the outputs of AI that complies with the ASA and ACM ethical practice standards will meet both responsibilities outlined in the *Toronto Declaration* and will also meet the common elements identified in syntheses of "ethical guidelines" like Ryan & Stahl (2020).

It is worth reiterating that privacy and confidentiality with respect to data are protected by the ethical computing and ethical statistics pracititioner. Whether or not that

individual is trained in either field, has a job with either in the title or scope of duties, or belongs to either professional organization, these responsibilities to utilize computing and statistical practices in an ethical manner still accrue. It is also well worth noting that "ethical AI" is much more than simply AI that maintains privacy and confidentiality. Specifically, both ethical computing and ethical statistical practices support developers, deployers, and users of AI in meeting their obligations to promote, protect, and respect fundamental and universal human rights.

3. Conclusions

Ethical AI is defined as AI that is developed, deployed, and utilized in accordance with the ethical practice standards for computing (ACM) and statistics and data science (ASA). This compliance will bring the practitioner into alignment with obligations articulated by the *Toronto* Declaration, and will also place the practitioner well within the "*frameworks to guide and inform dialogue and debate* around the non-technical implications of these technologies" (IEEE 2017, p. 3; emphasis added). Work that is carried out in compliance with the ACM and ASA ethical practice standards will enable practitioners, and leaders of teams practicing or utilizing the outputs of computing, statistics and data science, or AI to meet their obligations to promote and protect universal human rights. Those leading teams of cross- and multi-disciplinary workers can, and are charged by the ACM and ASA specifically to, enable their teams to do their work ethically when they engage with data, statistical practices, and AI. The specific support of the ethical practice standards of the ACM and ASA for the human rights outlined in the *Toronto Declaration* will help leaders recognize and accept responsibility for their and their team's compliance with these standards. Through this engagement, leaders can support institutionalizing ethical practices with data, statistics and data science, and AI. Recent books introduce ethical reasoning, and show how to reason with these guidelines specifically (Tractenberg 2022-A; 2022-B). Ethical reasoning can be deployed in the workplace using the ASA or ACM ethical practice standards, as well as other policy or guidance (e.g., United Nations Fundamental Principles of Statistics (2014), OECD Good Statistical Practices (2014), Principles and Practices for Federal Statistical Agencies and Recognized Statistical Units (2021), Data Ethics Tenets (OMB; 2020); Park & Tractenberg, 2023; Tractenberg & Park 2023).

State as well as private sector actors across diverse environments, whether practitioners or leaders, need to understand how ethical practice standards inform statistical and data science (ASA) and computational (ACM) choices. Although they do not need to have the technical expertise, leaders of teams utilizing computing, statistics and data science, and AI do need to understand the context in which technical decisions are made, and how ethical practice standards, together with relevant policies, enable all those involved to make these decisions ethically. The ASA and ACM ethical practice standards can be used to create and support an ethical workplace culture and environment for individuals and teams that does statistical, data science, and/or AI work.

References

Association for Computing Machinery (ACM). (2018) *Code of Ethics.* Downloaded from https://www.acm.org/about-acm/code-of-ethics on 12 October 2018.

American Statistical Association (ASA). (2022). *ASA Ethical Guidelines for Statistical Practice*-revised, downloaded from https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx on 30 April 2018/2 February 2022.

Balkin JM. (2016). Information fiduciaries and the first amendment. UC Davis Law Review, 49, 1183–1234.

Council of Europe (2018). Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries.

Gillikin J, Kopolow A & Schrimmer K. (2017). *Principles for the Development of a Professional Code of Ethics* [white paper]. National Association for Healthcare Quality. *SocArXiv* https://osf.io/preprints/socarxiv/dt5kr/

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Latonero M. (2018). Governing Artificial Intelligence: Upholding human rights and dignity. *Data & Society*. https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1, 501–507.

National Academies of Science, Engineering, and Medicine (NASEM). (2021). *Principles and Practices for a Federal Statistical Agency, Seventh Edition*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25885.

Novelli C, Taddeo M, & Floridi L. (2023). Accountability in artificial intelligence: what it is and how it works. *AI & Society*. https://doi.org/10.1007/s00146-023-01635-y

Office of Management and Budget. (2020). *Data ethics tenets*. Downloaded from https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf on 1 July 2020.

Organisation for Economic Co-operation and Development (OECD). (2019). Recommendation of the Council on Good Statistical Practice. C/M(2019)7 1394th

SESSION, agenda item 55 C(2019)28 (20 February 2019 adopted 13 March 2019). https://one.oecd.org/document/C(2019)28/en/pdf

Park J & Tractenberg RE. (in press-2023). How do ASA Ethical Guidelines for Statistical Practice Support U.S. Guidelines for Official Statistics? In, H. Doosti, (Ed.). *Ethical Statistics.* Cambridge, UK: Ethics International Press. Preprint available at: *StatArXiv* http://arxiv.org/abs/2309.07180

Ryan, M., & Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: Clarifying their content and normative implications. J*ournal of Information, Communication and Ethics in Society,* 19, 61–86.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017) Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, Version 2. IEEE. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine
Learning systems. (May 2018). https://www.accessnow.org/cms/assets/uploads/2018/05/Toronto-Declaration-D0V2.pdf.

Tractenberg RE. (2022). Ethical Reasoning for a Data-Centered World. (408 pages). Cambridge, UK: Ethics International.

Tractenberg RE. (2022). Ethical Practice of Statistics and Data Science. (678 pages). Cambridge, UK: Ethics International.

Tractenberg RE. (2023, May 6). Degrees of Freedom Analysis: A mixed method for theory building, decision making, and prediction. *SocArXiv*, osf.io/preprints/socarxiv/r5a7z

Tractenberg RE & Park J. (in press-2023). How does international guidance for statistical practice align with the ASA Ethical Guidelines? In, H. Doosti, (Ed.). Ethical Statistics. Cambridge, UK: Ethics International Press. Preprint available at: StatArXiv, https://arxiv.org/abs/2309.08713

United Nations Economic and Social Council. (2013). *Fundamental Principles of Official Statistics.* ECOSOC Res 2013/21, Substantive session of 2013, Agenda item 13 (c) E/RES/2013/21 (28 October 2013 adopted 24 July). https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf

United Nations Office of the High Commissioner for Human Rights. (1989). CCPR General Comment No. 18: Non-discrimination, 10 November 1989, available at: https://www.refworld.org/docid/453883fa8.html. accessed 26 October 2023

United Nations Office of the High Commissioner for Human Rights. (2017), Tackling Discrimination against Lesbian, Gay, Bi, Trans, & Intersex People Standards of Conduct for Business. https://www.unfe.org/wp-content/uploads/2017/09/UN-Standards-of-Conduct.pdf

APPENDIX

The Toronto Declaration
2018
ENGLISH https://www.torontodeclaration.org/declaration-text/english/

downloaded 4 july 2023

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 the discriminatory risks of machine learning systems. (**2**) We wish to complement this existing work by reaffirming the role of human rights law and standards in protecting individuals and groups from discrimination in any context. The human rights law and standards referenced in this Declaration provide solid foundations for developing ethical frameworks for machine learning, including provisions for accountability and means for remedy.

4. From policing, to welfare systems, to healthcare provision, to platforms for online discourse – to name a few examples – systems employing machine learning technologies can vastly and rapidly reinforce or change power structures on an unprecedented scale and with significant harm to human rights, notably the right to equality. There is a substantive and growing body of evidence to show that machine learning systems, which can be opaque and include unexplainable processes, can contribute to discriminatory or otherwise repressive practices if adopted and implemented without necessary safeguards.

5. States and private sector actors should promote the development and use of machine learning and related technologies where they help people exercise and enjoy their human rights. For example, in healthcare, machine learning systems could bring advances in diagnostics and treatments, while potentially making healthcare services more widely available and accessible. In relation to machine learning and artificial intelligence systems more broadly, states should promote the positive right to the enjoyment of developments in science and technology as an affirmation of economic, social and cultural rights. (**3**)

6. We focus in this Declaration on the right to equality and non-discrimination. There are numerous other human rights that may be adversely affected through the use and misuse of machine learning systems, including the right to privacy and data protection, the right to freedom of expression and association, to participation in cultural life, equality before the law, and access to effective remedy. Systems that make decisions and process data can also undermine economic, social, and cultural rights; for example, they can impact the provision of vital services, such as healthcare and education, and limit access to opportunities like employment.

7. While this Declaration is focused on machine learning technologies, many of the norms and principles included here are equally applicable to technologies housed under the broader term of artificial intelligence, as well as to related data systems.

The Declaration

- **Using the framework of international human rights law**
    - **The right to equality and non-discrimination**
    - **Preventing discrimination**
    - **Protecting the rights of all individuals and groups: promoting diversity and inclusion**
- **Duties of states: human rights obligations**
    - **State use of machine learning systems**
    - **Promoting equality**
    - **Holding private sector actors to account**
- **Responsibilities of private sector actors: human rights due diligence**
- **The right to an effective remedy**
- **Conclusion**
- **References**

Using the framework of international human rights law

8. States have obligations to promote, protect and respect human rights; private sector actors, including companies, have a responsibility to respect human rights at all times. We put forward this Declaration to affirm these obligations and responsibilities.

9. There are many discussions taking place now at supranational, state and regional level, in technology companies, at academic institutions, in civil society and beyond, focusing on the

ethics of artificial intelligence and how to make technology in this field human-centric. These issues must be analyzed through a human rights lens to assess current and future potential human rights harms created or facilitated by this technology, and to take concrete steps to address any risk of harm.

10. Human rights law is a universally ascribed system of values based on the rule of law. It provides established means to ensure that rights are upheld, including the rights to equality and non-discrimination. Its nature as a universally binding, actionable set of standards is particularly well-suited for borderless technologies. Human rights law sets standards and provides mechanisms to hold public and private sector actors accountable where they fail to fulfil their respective obligations and responsibilities to protect and respect rights. It also requires that everyone must be able to obtain effective remedy and redress where their rights have been denied or violated.

11. The risks that machine learning systems pose must be urgently examined and addressed at governmental level and by private sector actors who are conceiving, developing and deploying these systems. It is critical that potential harms are identified and addressed and that mechanisms are put in place to hold those responsible for harms to account. Government measures should be binding and adequate to protect and promote rights. Academic, legal and civil society experts should be able to meaningfully participate in these discussions, and critique and advise on the use of these technologies.

The right to equality and non-discrimination

12. This Declaration focuses on the right to equality and non-discrimination, a critical principle that underpins all human rights.

13. Discrimination is defined under international law as “any distinction, exclusion, restriction or preference which is based on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, and which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise by all persons, on an equal footing, of all rights and freedoms.” (**4**) This list is non-exhaustive as the United Nations High Commissioner for Human Rights has recognized the necessity of preventing discrimination against additional classes. (**5**)

Preventing discrimination

14. Governments have obligations and private sector actors have responsibilities to proactively prevent discrimination in order to comply with existing human rights law and standards. When prevention is not sufficient or satisfactory, and discrimination arises, a system should be interrogated and harms addressed immediately.

15. In employing new technologies, both state and private sector actors will likely need to find new ways to protect human rights, as new challenges to equality and representation of and impact on diverse individuals and groups arise.

16. Existing patterns of structural discrimination may be reproduced and aggravated in situations that are particular to these technologies – for example, machine learning system goals that create self-fulfilling markers of success and reinforce patterns of inequality, or issues arising from using non-representative or biased datasets.

17. All actors, public and private, must prevent and mitigate against discrimination risks in the design, development and application of machine learning technologies. They must also ensure that there are mechanisms allowing for access to effective remedy in place before deployment and throughout a system's lifecycle.

Protecting the rights of all individuals and groups: promoting diversity and inclusion

18. This Declaration underlines that inclusion, diversity and equity are key components of protecting and upholding the right to equality and non-discrimination. All must be considered in the development and deployment of machine learning systems in order to prevent discrimination, particularly against marginalised groups.

19. While the collection of data can help mitigate discrimination, there are some groups for whom collecting data on discrimination poses particular difficulty. Additional protections must extend to those groups, including protections for sensitive data.

20. Implicit and inadvertent bias through design creates another means for discrimination, where the conception, development and end use of machine learning systems is largely overseen by a particular sector of society. This technology is at present largely developed, applied and reviewed by companies based in certain countries and regions; the people behind the technology bring their own biases, and are likely to have limited input from diverse groups in terms of race, culture, gender, and socio-economic backgrounds.

21. Inclusion, diversity and equity entails the active participation of, and meaningful consultation with, a diverse community, including end users, during the design and application of machine learning systems, to help ensure that systems are created and used in ways that respect rights – particularly the rights of marginalised groups who are vulnerable to discrimination.

Duties of states: human rights obligations

22. States bear the primary duty to promote, protect, respect and fulfil human rights. Under international law, states must not engage in, or support discriminatory or otherwise rights-violating actions or practices when designing or implementing machine learning systems in a public context or through public-private partnerships.

23. States must adhere to relevant national and international laws and regulations that codify and implement human rights obligations protecting against discrimination and other related rights harms, for example data protection and privacy laws.

24. States have positive obligations to protect against discrimination by private sector actors and promote equality and other rights, including through binding laws.

25. The state obligations outlined in this section also apply to public use of machine learning in partnerships with private sector actors.

State use of machine learning systems

26. States must ensure that existing measures to prevent against discrimination and other rights harms are updated to take into account and address the risks posed by machine learning technologies.

27. Machine learning systems are increasingly being deployed or implemented by public authorities in areas that are fundamental to the exercise and enjoyment of human rights, rule of law, due process, freedom of expression, criminal justice, healthcare, access to social welfare benefits, and housing. While this technology may offer benefits in such contexts, there may also be a high risk of discriminatory or other rights-harming outcomes. It is critical that states provide meaningful opportunities for effective remediation and redress of harms where they do occur.

28. As confirmed by the Human Rights Committee, Article 26 of the International Covenant on Civil and Political Rights “prohibits discrimination in law or in fact in any field regulated and protected by public authorities”. (**6**) This is further set out in treaties dealing with specific forms of discrimination, in which states have committed to refrain from engaging in discrimination, and to ensure that public authorities and institutions “act in conformity with this obligation”. (**7**)

29. States must refrain altogether from using or requiring the private sector to use tools that discriminate, lead to discriminatory outcomes, or otherwise harm human rights.

30. **States must take the following steps to mitigate and reduce the harms of discrimination from machine learning in public sector systems:**

i. Identify risks

31. Any state deploying machine learning technologies must thoroughly investigate systems for discrimination and other rights risks prior to development or acquisition, where possible, prior to use, and on an ongoing basis throughout the lifecycle of the technologies, in the contexts in which they are deployed. This may include:

a) Conducting regular impact assessments prior to public procurement, during development, at regular milestones and throughout the deployment and use of machine learning systems to identify potential sources of discriminatory or other rights-harming outcomes – for example, in algorithmic model design, in oversight processes, or in data processing. (**8**)

b) Taking appropriate measures to mitigate risks identified through impact assessments – for example, mitigating inadvertent discrimination or underrepresentation in data or systems; conducting dynamic testing methods and pre-release trials; ensuring that potentially affected groups and field experts are included as actors with decision-making power in the design, testing and review phases; submitting systems for independent expert review where appropriate.

c) Subjecting systems to live, regular tests and audits; interrogating markers of success for bias and self-fulfilling feedback loops; and ensuring holistic independent reviews of systems in the context of human rights harms in a live environment.

d) Disclosing known limitations of the system in question – for example, noting measures of confidence, known failure scenarios and appropriate limitations of use.

ii. Ensure transparency and accountability

32. States must ensure and require accountability and maximum possible transparency around public sector use of machine learning systems. This must include explainability and intelligibility in the use of these technologies so that the impact on affected individuals and groups can be effectively scrutinised by independent entities, responsibilities established, and actors held to account. States should:

a) Publicly disclose where machine learning systems are used in the public sphere, provide information that explains in clear and accessible terms how automated and machine learning decision-making processes are reached, and document actions taken to identify, document and mitigate against discriminatory or other rights-harming impacts.

b) Enable independent analysis and oversight by using systems that are auditable.

c) Avoid using ‘black box systems’ that cannot be subjected to meaningful standards of accountability and transparency, and refrain from using these systems at all in high-risk contexts. (**9**)

iii. Enforce oversight

33. States must take steps to ensure public officials are aware of and sensitive to the risks of discrimination and other rights harms in machine learning systems. States should:

a) Proactively adopt diverse hiring practices and engage in consultations to assure diverse perspectives so that those involved in the design, implementation, and review of machine learning represent a range of backgrounds and identities.

b) Ensure that public bodies carry out training in human rights and data analysis for officials involved in the procurement, development, use and review of machine learning tools.

c) Create mechanisms for independent oversight, including by judicial authorities when necessary.

d) Ensure that machine learning-supported decisions meet international accepted standards for due process.

34. As research and development of machine learning systems is largely driven by the private sector, in practice states often rely on private contractors to design and implement these technologies in a public context. In such cases, states must not relinquish their own obligations around preventing discrimination and ensuring accountability and redress for human rights harms in the delivery of services.

35. Any state authority procuring machine learning technologies from the private sector should maintain relevant oversight and control over the use of the system, and require the third party to carry out human rights due diligence to identify, prevent and mitigate against discrimination and other human rights harms, and publicly account for their efforts in this regard.

Promoting equality

36. **States have a duty to take proactive measures to eliminate discrimination.** (**10**)

37. In the context of machine learning and wider technology developments, one of the most important priorities for states is to promote programs that increase diversity, inclusion and equity in the science, technology, engineering and mathematics sectors (commonly referred to as STEM fields). Such efforts do not serve as ends in themselves, though they may help mitigate against discriminatory outcomes. States should also invest in research into ways to mitigate human rights harms in machine learning systems.

Holding private sector actors to account

38. **International law clearly sets out the duty of states to protect human rights; this includes ensuring the right to non-discrimination by private sector actors.**

39. According to the UN Committee on Economic, Social and Cultural Rights, “States parties must therefore adopt measures, which should include legislation, to ensure that individuals and entities in the private sphere do not discriminate on prohibited grounds”. (**11**)

40. States should put in place regulation compliant with human rights law for oversight of the use of machine learning by the private sector in contexts that present risk of discriminatory or other rights-harming outcomes, recognising technical standards may be complementary to regulation. In addition, non-discrimination, data protection, privacy and other areas of law at national and regional levels may expand upon and reinforce international human rights obligations applicable to machine learning.

41. States must guarantee access to effective remedy for all individuals whose rights are violated or abused through use of these technologies.

Responsibilities of private sector actors: human rights due diligence

42. Private sector actors have a responsibility to respect human rights; this responsibility exists independently of state obligations. (**12**) As part of fulfilling this responsibility, private sector actors need to take ongoing proactive and reactive steps to ensure that they do not cause or contribute to human rights abuses – a process called ‘human rights due diligence’. (**13**)

43. Private sector actors that develop and deploy machine learning systems should follow a human rights due diligence framework to avoid fostering or entrenching discrimination and to respect human rights more broadly through the use of their systems.

44. **There are three core steps to the process of human rights due diligence:**

i. Identify potential discriminatory outcomes
ii. Take effective action to prevent and mitigate discrimination and track responses
iii. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems.

i. Identify potential discriminatory outcomes

45. During the development and deployment of any new machine learning technologies, non-state and private sector actors should assess the risk that the system will result in discrimination. The risk of discrimination and the harms will not be equal in all applications, and the actions required to address discrimination will depend on the context. Actors must be careful to identify not only direct discrimination, but also indirect forms of differential treatment which may appear neutral at face value, but lead to discrimination.

46. When mapping risks, private sector actors should take into account risks commonly associated with machine learning systems – for example, training systems on incomplete or

unrepresentative data, or datasets representing historic or systemic bias. Private actors should consult with relevant stakeholders in an inclusive manner, including affected groups, organizations that work on human rights, equality and discrimination, as well as independent human rights and machine learning experts.

ii. Take effective action to prevent and mitigate discrimination and track responses

47. After identifying human rights risks, the second step is to prevent those risks. For developers of machine learning systems, this requires:

a) Correcting for discrimination, both in the design of the model and the impact of the system and in deciding which training data to use.

b) Pursuing diversity, equity and other means of inclusion in machine learning development teams, with the aim of identifying bias by design and preventing inadvertent discrimination.

c) Submitting systems that have a significant risk of resulting in human rights abuses to independent third-party audits.

48. Where the risk of discrimination or other rights violations has been assessed to be too high or impossible to mitigate, private sector actors should not deploy a machine learning system in that context.

49. Another vital element of this step is for private sector actors to track their response to issues that emerge during implementation and over time, including evaluation of the effectiveness of responses. This requires regular, ongoing quality assurances checks and real-time auditing through design, testing and deployment stages to monitor a system for discriminatory impacts in context and situ, and to correct errors and harms as appropriate. This is particularly important given the risk of feedback loops that can exacerbate and entrench discriminatory outcomes.

iii. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems

50. Transparency is a key component of human rights due diligence, and involves "communication, providing a measure of transparency and accountability to individuals or groups who may be impacted and to other relevant stakeholders." (**14**)

51. Private sector actors that develop and implement machine learning systems should disclose the process of identifying risks, the risks that have been identified, and the concrete steps taken to prevent and mitigate identified human rights risks. This may include:

a) Disclosing information about the risks and specific instances of discrimination the company has identified, for example risks associated with the way a particular machine learning system is designed, or with the use of machine learning systems in particular contexts.

b) In instances where there is a risk of discrimination, publishing technical specification with details of the machine learning and its functions, including samples of the training data used and details of the source of data.

c) Establishing mechanisms to ensure that where discrimination has occurred through the use of a machine learning system, relevant parties, including affected individuals, are informed of the harms and how they can challenge a decision or outcome.

The right to an effective remedy

52. The right to justice is a vital element of international human rights law. Under international law, victims of human rights violations or abuses must have access to prompt and effective remedies, and those responsible for the violations must be held to account. (**15**)

53. Companies and private sector actors designing and implementing machine learning systems should take action to ensure individuals and groups have access to meaningful, effective remedy and redress. This may include, for example, creating clear, independent, visible processes for redress following adverse individual or societal effects, and designating roles in the entity responsible for the timely remedy of such issues subject to accessible and effective appeal and judicial review.

54. The use of machine learning systems where people's rights are at stake may pose challenges for ensuring the right to remedy. The opacity of some systems means individuals may be unaware how decisions which affect their rights were made, and whether the process was discriminatory. In some cases, the public body or private sector actors involved may itself be unable to explain the decision-making process.

55. The challenges are particularly acute when machine learning systems that recommend, make or enforce decisions are used within the justice system, the very institutions which are responsible for guaranteeing rights, including the right to access to effective remedy.

56. The measures already outlined around identifying, documenting, and responding to discrimination, and being transparent and accountable about these efforts, will help states to ensure that individuals have access to effective remedies. In addition, states should:

a) Ensure that if machine learning systems are to be deployed in the public sector, use is carried out in line with standards of due process.

b) Act cautiously on the use of machine learning systems in justice sector given the risks to fair trial and litigants' rights. (**16**)

c) Outline clear lines of accountability for the development and implementation of machine learning systems and clarify which bodies or individuals are legally responsible for decisions made through the use of such systems.

d) Provide effective remedies to victims of discriminatory harms linked to machine learning systems used by public or private bodies, including reparation that, where appropriate, can involve compensation, sanctions against those responsible, and guarantees of non-repetition. This may be possible using existing laws and regulations or may require developing new ones.

Conclusion

57. The signatories of this Declaration call for public and private sector actors to uphold their obligations and responsibilities under human rights laws and standards to avoid discrimination in the use of machine learning systems where possible. Where discrimination arises, measures to deliver the right to effective remedy must be in place.

58. We call on states and private sector actors to work together and play an active and committed role in protecting individuals and groups from discrimination. When creating and deploying machine learning systems, they must take meaningful measures to promote accountability and human rights, including, but not limited to, the right to equality and non-discrimination, as per their obligations and responsibilities under international human rights law and standards.

59. Technological advances must not undermine our human rights. We are at a crossroads where those with the power must act now to protect human rights, and help safeguard the rights that we are all entitled to now, and for future generations.

This Declaration was published on 16 May 2018 by Amnesty International and Access Now, and launched at RightsCon 2018 in Toronto, Canada.

References (sic)

1. UN Human Rights Committee, **Vienna Declaration and Programme of Action**, 1993
2. For example, see the FAT/ML **Principles for Accountable Algorithms and a Social Impact Statement for Algorithms**; IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, **Ethically Aligned Design**; **The Montreal Declaration** for a Responsible Development of Artificial Intelligence; The **Asilomar AI Principles** developed by the Future of Life Institute.

3. **The International Covenant on Economic, Social and Cultural Rights** (ICESCR), Article 15
4. United Nations Human Rights Committee, General comment No. 18, UN Doc. RI/GEN/1/Rev.9 Vol. I (1989), para. 7
5. UN OHCHR, **Tackling Discrimination against Lesbian, Gay, Bi, Trans, & Intersex People Standards of Conduct for Business**
6. United Nations Human Rights Committee, General comment No. 18 (1989), para. 12
7. For example, **Convention on the Elimination of All Forms of Racial Discrimination**, Article 2 (a), and Convention on the Elimination of All Forms of Discrimination against Women, Article 2(d).
8. The AI Now Institute has outlined a practical framework for **algorithmic impact assessments by public agencies**. Article 35 of the EU's General Data Protection Regulation (GDPR) sets out a requirement to carry out a Data Protection Impact Assessment (DPIA); in addition, Article 25 of the GDPR requires data protection principles to be applied by design and by default from the conception phase of a product, service or service and through its lifecycle.
9. The AI Now Institute at New York University, **AI Now 2017 Report**, 2017
10. The UN Committee on Economic, Social and Cultural Rights affirms that in addition to refraining from discriminatory actions, "State parties should take concrete, deliberate and targeted measures to ensure that discrimination in the exercise of Covenant rights is eliminated." – UN Committee on Economic, Social and Cultural Rights, General Comment 20, UN Doc. E/C.12/GC/20 (2009) para. 36
11. UN Committee on Economic, Social and Cultural Rights, General Comment 20, UN Doc. E/C.12/GC/20 (2009) para. 11
12. **UN Guiding Principles on Business and Human Rights** and additional supporting documents
13. Council of Europe Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the **roles and responsibilities of internet intermediaries**
14. **UN Guiding Principles on Business and Human Rights**, Principle 21
15. For example, see: Universal Declaration of Human Rights, Article 8; International Covenant on Civil and Political Rights, Article 2 (3); International Covenant on Economic, Social and Cultural Rights, Article 2; Committee on Economic, Social and Cultural Rights, General Comment No. 3: The Nature of States Parties' Obligations, UN Doc. E/1991/23 (1990) Article 2 Para. 1 of the Covenant; International Convention on the Elimination of All Forms of Racial Discrimination, Article 6; Convention on the Elimination of All Forms of Discrimination against Women and UN Committee on Economic, Social and Cultural Rights (CESCR), Article 2, **General Comment No. 9**: The domestic application of the Covenant, E/C.12/1998/24 (1998)
16. For example, see: Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner for ProPublica, **Machine Bias**, 2016

ACM Code of Ethics and Professional Conduct

Adopted by ACM Council 6/22/18.

Preamble

Computing professionals' actions change the world. To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good. The ACM Code of Ethics and Professional Conduct ("the Code") expresses the conscience of the profession.

The Code is designed to inspire and guide the ethical conduct of all computing professionals, including current and aspiring practitioners, instructors, students, influencers, and anyone who uses computing technology in an impactful way. Additionally, the Code serves as a basis for remediation when violations occur. The Code includes principles formulated as statements of responsibility, based on the understanding that the public good is always the primary consideration. Each principle is supplemented by guidelines, which provide explanations to assist computing professionals in understanding and applying the principle.

Section 1 outlines fundamental ethical principles that form the basis for the remainder of the Code. Section 2 addresses additional, more specific considerations of professional responsibility. Section 3 guides individuals who have a leadership role, whether in the workplace or in a volunteer professional capacity. Commitment to ethical conduct is required of every ACM member, and principles involving compliance with the Code are given in Section 4.

The Code as a whole is concerned with how fundamental ethical principles apply to a computing professional's conduct. The Code is not an algorithm for solving ethical problems; rather it serves as a basis for ethical decision-making. When thinking through a particular issue, a computing professional may find that multiple principles should be taken into account, and that different principles will have different relevance to the issue. Questions related to these kinds of issues can best be answered by thoughtful consideration of the fundamental ethical principles, understanding that the public good is the paramount consideration. The entire computing profession benefits when the ethical decision-making process is accountable to and transparent to all stakeholders. Open discussions about ethical issues promote this accountability and transparency.

1. GENERAL ETHICAL PRINCIPLES.

A computing professional should…

1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.

This principle, which concerns the quality of life of all people, affirms an obligation of computing professionals, both individually and collectively, to use their skills for the benefit of society, its members, and the environment surrounding them. This obligation includes promoting fundamental human rights and protecting each individual's right to autonomy. An essential aim of computing professionals is to minimize negative consequences of computing, including threats to health, safety, personal security, and privacy. When the interests of multiple groups conflict, the needs of those less advantaged should be given increased attention and priority.

Computing professionals should consider whether the results of their efforts will respect diversity, will be used in socially responsible ways, will meet social needs, and will be broadly accessible. They are encouraged to actively contribute to society by engaging in pro bono or volunteer work that benefits the public good.

In addition to a safe social environment, human well-being requires a safe natural environment. Therefore, computing professionals should promote environmental sustainability both locally and globally.

1.2 Avoid harm.

In this document, “harm” means negative consequences, especially when those consequences are significant and unjust. Examples of harm include unjustified physical or mental injury, unjustified destruction or disclosure of information, and unjustified damage to property, reputation, and the environment. This list is not exhaustive.

Well-intended actions, including those that accomplish assigned duties, may lead to harm. When that harm is unintended, those responsible are obliged to undo or mitigate the harm as much as possible. Avoiding harm begins with careful consideration of potential impacts on all those affected by decisions. When harm is an intentional part of the system, those responsible are obligated to ensure that the harm is ethically justified. In either case, ensure that all harm is minimized.

To minimize the possibility of indirectly or unintentionally harming others, computing professionals should follow generally accepted best practices unless there is a compelling ethical reason to do otherwise. Additionally, the consequences of data aggregation and emergent properties of systems should be carefully analyzed. Those involved with pervasive or infrastructure systems should also consider Principle 3.7.

A computing professional has an additional obligation to report any signs of system risks that might result in harm. If leaders do not act to curtail or mitigate such risks, it may be necessary to “blow the whistle” to reduce potential harm. However, capricious or misguided reporting of risks can itself be harmful. Before reporting risks, a computing professional should carefully assess relevant aspects of the situation.

1.3 Be honest and trustworthy.

Honesty is an essential component of trustworthiness. A computing professional should be transparent and provide full disclosure of all pertinent system capabilities, limitations, and potential problems to the appropriate parties. Making deliberately false or misleading claims, fabricating or falsifying data, offering or accepting bribes, and other dishonest conduct are violations of the Code.

Computing professionals should be honest about their qualifications, and about any limitations in their competence to complete a task. Computing professionals should be forthright about any

circumstances that might lead to either real or perceived conflicts of interest or otherwise tend to undermine the independence of their judgment. Furthermore, commitments should be honored.

Computing professionals should not misrepresent an organization's policies or procedures, and should not speak on behalf of an organization unless authorized to do so.

1.4 Be fair and take action not to discriminate.

The values of equality, tolerance, respect for others, and justice govern this principle. Fairness requires that even careful decision processes provide some avenue for redress of grievances.

Computing professionals should foster fair participation of all people, including those of underrepresented groups. Prejudicial discrimination on the basis of age, color, disability, ethnicity, family status, gender identity, labor union membership, military status, nationality, race, religion or belief, sex, sexual orientation, or any other inappropriate factor is an explicit violation of the Code. Harassment, including sexual harassment, bullying, and other abuses of power and authority, is a form of discrimination that, amongst other harms, limits fair access to the virtual and physical spaces where such harassment takes place.

The use of information and technology may cause new, or enhance existing, inequities. Technologies and practices should be as inclusive and accessible as possible and computing professionals should take action to avoid creating systems or technologies that disenfranchise or oppress people. Failure to design for inclusiveness and accessibility may constitute unfair discrimination.

1.5 Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.

Developing new ideas, inventions, creative works, and computing artifacts creates value for society, and those who expend this effort should expect to gain value from their work. Computing professionals should therefore credit the creators of ideas, inventions, work, and artifacts, and respect copyrights, patents, trade secrets, license agreements, and other methods of protecting authors' works.

Both custom and the law recognize that some exceptions to a creator's control of a work are necessary for the public good. Computing professionals should not unduly oppose reasonable uses of their intellectual works. Efforts to help others by contributing time and energy to projects that help society illustrate a positive aspect of this principle. Such efforts include free and open source software and work put into the public domain. Computing professionals should not claim private ownership of work that they or others have shared as public resources.

1.6 Respect privacy.

The responsibility of respecting privacy applies to computing professionals in a particularly profound way. Technology enables the collection, monitoring, and exchange of personal information quickly, inexpensively, and often without the knowledge of the people affected. Therefore, a computing professional should become conversant in the various definitions and forms of privacy and should understand the rights and responsibilities associated with the collection and use of personal information.

Computing professionals should only use personal information for legitimate ends and without violating the rights of individuals and groups. This requires taking precautions to prevent re-identification of anonymized data or unauthorized data collection, ensuring the accuracy of data, understanding the provenance of the data, and protecting it from unauthorized access and accidental disclosure. Computing professionals should establish transparent policies and procedures that allow individuals to understand what data is being collected and how it is being used, to give informed consent for automatic data collection, and to review, obtain, correct inaccuracies in, and delete their personal data.

Only the minimum amount of personal information necessary should be collected in a system. The retention and disposal periods for that information should be clearly defined, enforced, and communicated to data subjects. Personal information gathered for a specific purpose should not be used for other purposes without the person's consent. Merged data collections can compromise privacy features present in the original collections. Therefore, computing professionals should take special care for privacy when merging data collections.

1.7 Honor confidentiality.

Computing professionals are often entrusted with confidential information such as trade secrets, client data, nonpublic business strategies, financial information, research data, pre-publication scholarly articles, and patent applications. Computing professionals should protect confidentiality except in cases where it is evidence of the violation of law, of organizational regulations, or of the Code. In these cases, the nature or contents of that information should not be disclosed except to appropriate authorities. A computing professional should consider thoughtfully whether such disclosures are consistent with the Code.

2. PROFESSIONAL RESPONSIBILITIES.

A computing professional should…

2.1 Strive to achieve high quality in both the processes and products of professional work.

Computing professionals should insist on and support high quality work from themselves and from colleagues. The dignity of employers, employees, colleagues, clients, users, and anyone else affected either directly or indirectly by the work should be respected throughout the process. Computing professionals should respect the right of those involved to transparent

communication about the project. Professionals should be cognizant of any serious negative consequences affecting any stakeholder that may result from poor quality work and should resist inducements to neglect this responsibility.

2.2 Maintain high standards of professional competence, conduct, and ethical practice.

High quality computing depends on individuals and teams who take personal and group responsibility for acquiring and maintaining professional competence. Professional competence starts with technical knowledge and with awareness of the social context in which their work may be deployed. Professional competence also requires skill in communication, in reflective analysis, and in recognizing and navigating ethical challenges. Upgrading skills should be an ongoing process and might include independent study, attending conferences or seminars, and other informal or formal education. Professional organizations and employers should encourage and facilitate these activities.

2.3 Know and respect existing rules pertaining to professional work.

“Rules” here include local, regional, national, and international laws and regulations, as well as any policies and procedures of the organizations to which the professional belongs. Computing professionals must abide by these rules unless there is a compelling ethical justification to do otherwise. Rules that are judged unethical should be challenged. A rule may be unethical when it has an inadequate moral basis or causes recognizable harm. A computing professional should consider challenging the rule through existing channels before violating the rule. A computing professional who decides to violate a rule because it is unethical, or for any other reason, must consider potential consequences and accept responsibility for that action.

2.4 Accept and provide appropriate professional review.

High quality professional work in computing depends on professional review at all stages. Whenever appropriate, computing professionals should seek and utilize peer and stakeholder review. Computing professionals should also provide constructive, critical reviews of others’ work.

2.5 Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks.

Computing professionals are in a position of trust, and therefore have a special responsibility to provide objective, credible evaluations and testimony to employers, employees, clients, users, and the public. Computing professionals should strive to be perceptive, thorough, and objective when evaluating, recommending, and presenting system descriptions and alternatives. Extraordinary care should be taken to identify and mitigate potential risks in machine learning systems. A system for which future risks cannot be reliably predicted requires frequent reassessment of risk as the system evolves in use, or it should not be deployed. Any issues that might result in major risk must be reported to appropriate parties.

2.6 Perform work only in areas of competence.

A computing professional is responsible for evaluating potential work assignments. This includes evaluating the work's feasibility and advisability, and making a judgment about whether the work assignment is within the professional's areas of competence. If at any time before or during the work assignment the professional identifies a lack of a necessary expertise, they must disclose this to the employer or client. The client or employer may decide to pursue the assignment with the professional after additional time to acquire the necessary competencies, to pursue the assignment with someone else who has the required expertise, or to forgo the assignment. A computing professional's ethical judgment should be the final guide in deciding whether to work on the assignment.

2.7 Foster public awareness and understanding of computing, related technologies, and their consequences.

As appropriate to the context and one's abilities, computing professionals should share technical knowledge with the public, foster awareness of computing, and encourage understanding of computing. These communications with the public should be clear, respectful, and welcoming. Important issues include the impacts of computer systems, their limitations, their vulnerabilities, and the opportunities that they present. Additionally, a computing professional should respectfully address inaccurate or misleading information related to computing.

2.8 Access computing and communication resources only when authorized or when compelled by the public good.

Individuals and organizations have the right to restrict access to their systems and data so long as the restrictions are consistent with other principles in the Code. Consequently, computing professionals should not access another's computer system, software, or data without a reasonable belief that such an action would be authorized or a compelling belief that it is consistent with the public good. A system being publicly accessible is not sufficient grounds on its own to imply authorization. Under exceptional circumstances a computing professional may use unauthorized access to disrupt or inhibit the functioning of malicious systems; extraordinary precautions must be taken in these instances to avoid harm to others.

2.9 Design and implement systems that are robustly and usably secure.

Breaches of computer security cause harm. Robust security should be a primary consideration when designing and implementing systems. Computing professionals should perform due diligence to ensure the system functions as intended, and take appropriate action to secure resources against accidental and intentional misuse, modification, and denial of service. As threats can arise and change after a system is deployed, computing professionals should integrate mitigation techniques and policies, such as monitoring, patching, and vulnerability reporting. Computing professionals should also take steps to ensure parties affected by data

breaches are notified in a timely and clear manner, providing appropriate guidance and remediation.

To ensure the system achieves its intended purpose, security features should be designed to be as intuitive and easy to use as possible. Computing professionals should discourage security precautions that are too confusing, are situationally inappropriate, or otherwise inhibit legitimate use.

In cases where misuse or harm are predictable or unavoidable, the best option may be to not implement the system.

3. PROFESSIONAL LEADERSHIP PRINCIPLES.

Leadership may either be a formal designation or arise informally from influence over others. In this section, “leader” means any member of an organization or group who has influence, educational responsibilities, or managerial responsibilities. While these principles apply to all computing professionals, leaders bear a heightened responsibility to uphold and promote them, both within and through their organizations.

A computing professional, especially one acting as a leader, should…

3.1 Ensure that the public good is the central concern during all professional computing work.

People—including users, customers, colleagues, and others affected directly or indirectly—should always be the central concern in computing. The public good should always be an explicit consideration when evaluating tasks associated with research, requirements analysis, design, implementation, testing, validation, deployment, maintenance, retirement, and disposal. Computing professionals should keep this focus no matter which methodologies or techniques they use in their practice.

3.2 Articulate, encourage acceptance of, and evaluate fulfillment of social responsibilities by members of the organization or group.

Technical organizations and groups affect broader society, and their leaders should accept the associated responsibilities. Organizations—through procedures and attitudes oriented toward quality, transparency, and the welfare of society—reduce harm to the public and raise awareness of the influence of technology in our lives. Therefore, leaders should encourage full participation of computing professionals in meeting relevant social responsibilities and discourage tendencies to do otherwise.

3.3 Manage personnel and resources to enhance the quality of working life.

Leaders should ensure that they enhance, not degrade, the quality of working life. Leaders should consider the personal and professional development, accessibility requirements,

physical safety, psychological well-being, and human dignity of all workers. Appropriate human-computer ergonomic standards should be used in the workplace.

3.4 Articulate, apply, and support policies and processes that reflect the principles of the Code.

Leaders should pursue clearly defined organizational policies that are consistent with the Code and effectively communicate them to relevant stakeholders. In addition, leaders should encourage and reward compliance with those policies, and take appropriate action when policies are violated. Designing or implementing processes that deliberately or negligently violate, or tend to enable the violation of, the Code's principles is ethically unacceptable.

3.5 Create opportunities for members of the organization or group to grow as professionals.

Educational opportunities are essential for all organization and group members. Leaders should ensure that opportunities are available to computing professionals to help them improve their knowledge and skills in professionalism, in the practice of ethics, and in their technical specialties. These opportunities should include experiences that familiarize computing professionals with the consequences and limitations of particular types of systems. Computing professionals should be fully aware of the dangers of oversimplified approaches, the improbability of anticipating every possible operating condition, the inevitability of software errors, the interactions of systems and their contexts, and other issues related to the complexity of their profession—and thus be confident in taking on responsibilities for the work that they do.

3.6 Use care when modifying or retiring systems.

Interface changes, the removal of features, and even software updates have an impact on the productivity of users and the quality of their work. Leaders should take care when changing or discontinuing support for system features on which people still depend. Leaders should thoroughly investigate viable alternatives to removing support for a legacy system. If these alternatives are unacceptably risky or impractical, the developer should assist stakeholders' graceful migration from the system to an alternative. Users should be notified of the risks of continued use of the unsupported system long before support ends. Computing professionals should assist system users in monitoring the operational viability of their computing systems, and help them understand that timely replacement of inappropriate or outdated features or entire systems may be needed.

3.7 Recognize and take special care of systems that become integrated into the infrastructure of society.

Even the simplest computer systems have the potential to impact all aspects of society when integrated with everyday activities such as commerce, travel, government, healthcare, and education. When organizations and groups develop systems that become an important part of

the infrastructure of society, their leaders have an added responsibility to be good stewards of these systems. Part of that stewardship requires establishing policies for fair system access, including for those who may have been excluded. That stewardship also requires that computing professionals monitor the level of integration of their systems into the infrastructure of society. As the level of adoption changes, the ethical responsibilities of the organization or group are likely to change as well. Continual monitoring of how society is using a system will allow the organization or group to remain consistent with their ethical obligations outlined in the Code. When appropriate standards of care do not exist, computing professionals have a duty to ensure they are developed.

4. COMPLIANCE WITH THE CODE.

A computing professional should…

4.1 Uphold, promote, and respect the principles of the Code.

The future of computing depends on both technical and ethical excellence. Computing professionals should adhere to the principles of the Code and contribute to improving them. Computing professionals who recognize breaches of the Code should take actions to resolve the ethical issues they recognize, including, when reasonable, expressing their concern to the person or persons thought to be violating the Code.

4.2 Treat violations of the Code as inconsistent with membership in the ACM.

Each ACM member should encourage and support adherence by all computing professionals regardless of ACM membership. ACM members who recognize a breach of the Code should consider reporting the violation to the ACM, which may result in remedial action as specified in the ACM's Code of Ethics and Professional Conduct Enforcement Policy.

The Code and guidelines were developed by the ACM Code 2018 Task Force: Executive Committee Don Gotterbarn (Chair), Bo Brinkman, Catherine Flick, Michael S Kirkpatrick, Keith Miller, Kate Varansky, and Marty J Wolf. Members: Eve Anderson, Ron Anderson, Amy Bruckman, Karla Carter, Michael Davis, Penny Duquenoy, Jeremy Epstein, Kai Kimppa, Lorraine Kisselburgh, Shrawan Kumar, Andrew McGettrick, Natasa Milic-Frayling, Denise Oram, Simon Rogerson, David Shama, Janice Sipior, Eugene Spafford, and Les Waguespack. The Task Force was organized by the ACM Committee on Professional Ethics. Significant contributions to the Code were also made by the broader international ACM membership. This Code and its guidelines were adopted by the ACM Council on June 22nd, 2018.

Ethical Guidelines for Statistical Practice
Prepared by the Committee on Professional Ethics
of the American Statistical Association
February 2022

PURPOSE OF THE GUIDELINES:

The American Statistical Association's Ethical Guidelines for Statistical Practice are intended to help statistical practitioners make decisions ethically. In these Guidelines, “statistical practice” includes activities such as: designing the collection of, summarizing, processing, analyzing, interpreting, or presenting, data; as well as model or algorithm development and deployment. Throughout these Guidelines, the term "statistical practitioner" includes all those who engage in statistical practice, regardless of job title, profession, level, or field of degree. The Guidelines are intended for individuals, but these principles are also relevant to organizations that engage in statistical practice.

The Ethical Guidelines aim to promote accountability by informing those who rely on any aspects of statistical practice of the standards that they should expect. Society benefits from informed judgments supported by ethical statistical practice. All statistical practitioners are expected to follow these Guidelines and to encourage others to do the same.

In some situations, Guideline principles may require balancing of competing interests. If an unexpected ethical challenge arises, the ethical practitioner seeks guidance, not exceptions, in the Guidelines. To justify unethical behaviors, or to exploit gaps in the Guidelines, is unprofessional, and inconsistent with these Guidelines.

PRINCIPLE A: Professional Integrity and Accountability

Professional integrity and accountability require taking responsibility for one’s work. Ethical statistical practice supports valid and prudent decision making with appropriate methodology. The ethical statistical practitioner represents their capabilities and activities honestly, and treats others with respect.

The ethical statistical practitioner:

1. Takes responsibility for evaluating potential tasks, assessing whether they have (or can attain) sufficient competence to execute each task, and that the work and timeline are feasible. Does not solicit or deliver work for which they are not qualified, or that they would not be willing to have peer reviewed.
2. Uses methodology and data that are valid, relevant, and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results.
3. Does not knowingly conduct statistical practices that exploit vulnerable populations or create or perpetuate unfair outcomes.
4. Opposes efforts to predetermine or influence the results of statistical practices, and resists pressure to selectively interpret data.

5. Accepts full responsibility for their own work; does not take credit for the work of others; and gives credit to those who contribute. Respects and acknowledges the intellectual property of others.
6. Strives to follow, and encourages all collaborators to follow, an established protocol for authorship. Advocates for recognition commensurate with each person's contribution to the work. Recognizes that inclusion as an author does imply, while acknowledgement may imply, endorsement of the work.
7. Discloses conflicts of interest, financial and otherwise, and manages or resolves them according to established policies, regulations, and laws.
8. Promotes the dignity and fair treatment of all people. Neither engages in nor condones discrimination based on personal characteristics. Respects personal boundaries in interactions and avoids harassment including sexual harassment, bullying, and other abuses of power or authority.
9. Takes appropriate action when aware of deviations from these Guidelines by others.
10. Acquires and maintains competence through upgrading of skills as needed to maintain a high standard of practice.
11. Follows applicable policies, regulations, and laws relating to their professional work, unless there is a compelling ethical justification to do otherwise.
12. Upholds, respects, and promotes these Guidelines. Those who teach, train, or mentor in statistical practice have a special obligation to promote behavior that is consistent with these Guidelines.

PRINCIPLE B: Integrity of Data and Methods

The ethical statistical practitioner seeks to understand and mitigate known or suspected limitations, defects, or biases in the data or methods and communicates potential impacts on the interpretation, conclusions, recommendations, decisions, or other results of statistical practices.

The ethical statistical practitioner:

1. Communicates data sources and fitness for use, including data generation and collection processes and known biases. Discloses and manages any conflicts of interest relating to the data sources. Communicates data processing and transformation procedures, including missing data handling.
2. Is transparent about assumptions made in the execution and interpretation of statistical practices including methods used, limitations, possible sources of error, and algorithmic biases. Conveys results or applications of statistical practices in ways that are honest and meaningful.
3. Communicates the stated purpose and the intended use of statistical practices. Is transparent regarding a priori versus post hoc objectives and planned versus unplanned statistical practices. Discloses when multiple comparisons are conducted, and any relevant adjustments.
4. Meets obligations to share the data used in the statistical practices, for example, for peer review and replication, as allowable. Respects expectations of data contributors when using or sharing data. Exercises due caution to protect proprietary and confidential data, including all data that might inappropriately harm data subjects.

5. Strives to promptly correct substantive errors discovered after publication or implementation. As appropriate, disseminates the correction publicly and/or to others relying on the results.
6. For models and algorithms designed to inform or implement decisions repeatedly, develops and/or implements plans to validate assumptions and assess performance over time, as needed. Considers criteria and mitigation plans for model or algorithm failure and retirement.
7. Explores and describes the effect of variation in human characteristics and groups on statistical practice when feasible and relevant.

PRINCIPLE C: Responsibilities to Stakeholders

Those who fund, contribute to, use, or are affected by statistical practices are considered stakeholders. The ethical statistical practitioner respects the interests of stakeholders while practicing in compliance with these Guidelines.

The ethical statistical practitioner:

1. Seeks to establish what stakeholders hope to obtain from any specific project. Strives to obtain sufficient subject-matter knowledge to conduct meaningful and relevant statistical practice.
2. Regardless of personal or institutional interests or external pressures, does not use statistical practices to mislead any stakeholder.
3. Uses practices appropriate to exploratory and confirmatory phases of a project, differentiating findings from each so the stakeholders can understand and apply the results.
4. Informs stakeholders of the potential limitations on use and re-use of statistical practices in different contexts and offers guidance and alternatives, where appropriate, about scope, cost, and precision considerations that affect the utility of the statistical practice.
5. Explains any expected adverse consequences from failing to follow through on an agreed-upon sampling or analytic plan.
6. Strives to make new methodological knowledge widely available to provide benefits to society at large. Presents relevant findings, when possible, to advance public knowledge.
7. Understands and conforms to confidentiality requirements for data collection, release, and dissemination and any restrictions on its use established by the data provider (to the extent legally required). Protects the use and disclosure of data accordingly. Safeguards privileged information of the employer, client, or funder.
8. Prioritizes both scientific integrity and the principles outlined in these Guidelines when interests are in conflict.

PRINCIPLE D: Responsibilities to Research Subjects, Data Subjects, or those directly affected by statistical practices

The ethical statistical practitioner does not misuse or condone the misuse of data. They protect and respect the rights and interests of human and animal subjects. These responsibilities extend to those who will be directly affected by statistical practices.

The ethical statistical practitioner:

1. Keeps informed about and adheres to applicable rules, approvals, and guidelines for the protection and welfare of human and animal subjects. Knows when work requires ethical review and oversight.[3]
2. Makes informed recommendations for sample size and statistical practice methodology in order to avoid the use of excessive or inadequate numbers of subjects and excessive risk to subjects
3. For animal studies, seeks to leverage statistical practice to reduce the number of animals used, refine experiments to increase the humane treatment of animals, and replace animal use where possible.
4. Protects people's privacy and the confidentiality of data concerning them, whether obtained from the individuals directly, other persons, or existing records. Knows and adheres to applicable rules, consents, and guidelines to protect private information.
5. Uses data only as permitted by data subjects' consent when applicable or considering their interests and welfare when consent is not required. This includes primary and secondary uses, use of repurposed data, sharing data, and linking data with additional data sets.
6. Considers the impact of statistical practice on society, groups, and individuals. Recognizes that statistical practice could adversely affect groups or the public perception of groups, including marginalized groups. Considers approaches to minimize negative impacts in applications or in framing results in reporting.
7. Refrains from collecting or using more data than is necessary. Uses confidential information only when permitted and only to the extent necessary. Seeks to minimize the risk of re-identification when sharing de-identified data or results where there is an expectation of confidentiality. Explains any impact of de-identification on accuracy of results.
8. To maximize contributions of data subjects, considers how best to use available data sources for exploration, training, testing, validation, or replication as needed for the application. The ethical statistical practitioner appropriately discloses how the data is used for these purposes and any limitations.
9. Knows the legal limitations on privacy and confidentiality assurances and does not over-promise or assume legal privacy and confidentiality protections where they may not apply.
10. Understands the provenance of the data, including origins, revisions, and any restrictions on usage, and fitness for use prior to conducting statistical practices.
11. Does not conduct statistical practice that could reasonably be interpreted by subjects as sanctioning a violation of their rights. Seeks to use statistical practices to promote the just and impartial treatment of all individuals.

PRINCIPLE E: Responsibilities to members of multidisciplinary teams

Statistical practice is often conducted in teams made up of professionals with different professional standards. The statistical practitioner must know how to work ethically in this environment.

The ethical statistical practitioner:

[3] Examples of ethical review and oversight include an Institutional Review Board (IRB), an Institutional Animal Care and Use Committee (IACUC), or a compliance assessment.

1. Recognizes and respects that other professions may have different ethical standards and obligations. Dissonance in ethics may still arise even if all members feel that they are working towards the same goal. It is essential to have a respectful exchange of views.
2. Prioritizes these Guidelines for the conduct of statistical practice in cases where ethical guidelines conflict.
3. Ensures that all communications regarding statistical practices are consistent with these Guidelines. Promotes transparency in all statistical practices.
4. Avoids compromising validity for expediency. Regardless of pressure on or within the team, does not use inappropriate statistical practices.

PRINCIPLE F: Responsibilities to Fellow Statistical Practitioners and the Profession

Statistical practices occur in a wide range of contexts. Irrespective of job title and training, those who practice statistics have a responsibility to treat statistical practitioners, and the profession, with respect. Responsibilities to other practitioners and the profession include honest communication and engagement that can strengthen the work of others and the profession.

The ethical statistical practitioner:

1. Recognizes that statistical practitioners may have different expertise and experiences, which may lead to divergent judgments about statistical practices and results. Constructive discourse with mutual respect focuses on scientific principles and methodology and not personal attributes.
2. Helps strengthen, and does not undermine, the work of others through appropriate peer review or consultation. Provides feedback or advice that is impartial, constructive, and objective.
3. Takes full responsibility for their contributions as instructors, mentors, and supervisors of statistical practice by ensuring their best teaching and advising -- regardless of an academic or non-academic setting -- to ensure that developing practitioners are guided effectively as they learn and grow in their careers.
4. Promotes reproducibility and replication, whether results are “significant” or not, by sharing data, methods, and documentation to the extent possible.
5. Serves as an ambassador for statistical practice by promoting thoughtful choices about data acquisition, analytic procedures, and data structures among non-practitioners and students. Instills appreciation for the concepts and methods of statistical practice.

PRINCIPLE G: Responsibilities of Leaders, Supervisors, and Mentors in Statistical Practice

Statistical practitioners leading, supervising, and/or mentoring people in statistical practice have specific obligations to follow and promote these Ethical Guidelines. Their support for – and insistence on – ethical statistical practice are essential for the integrity of the practice and profession of statistics as well as the practitioners themselves.

Those leading, supervising, or mentoring statistical practitioners are expected to:

1. Ensure appropriate statistical practice that is consistent with these Guidelines. Protect the statistical practitioners who comply with these Guidelines, and advocate for a culture that supports ethical statistical practice.
2. Promote a respectful, safe, and productive work environment. Encourage constructive engagement to improve statistical practice.
3. Identify and/or create opportunities for team members/mentees to develop professionally and maintain their proficiency.
4. Advocate for appropriate, timely, inclusion and participation of statistical practitioners as contributors/collaborators. Promote appropriate recognition of the contributions of statistical practitioners, including authorship if applicable.
5. Establish a culture that values validation of assumptions, and assessment of model/algorithm performance over time and across relevant subgroups, as needed. Communicate with relevant stakeholders regarding model or algorithm maintenance, failure, or actual or proposed modifications.

PRINCIPLE H: Responsibilities Regarding Potential Misconduct

The ethical statistical practitioner understands that questions may arise concerning potential misconduct related to statistical, scientific, or professional practice. At times, a practitioner may accuse someone of misconduct, or be accused by others. At other times, a practitioner may be involved in the investigation of others' behavior. Allegations of misconduct may arise within different institutions with different standards and potentially different outcomes. The elements that follow relate specifically to allegations of statistical, scientific, and professional misconduct.

The ethical statistical practitioner:

1. Knows the definitions of, and procedures relating to, misconduct in their institutional setting. Seeks to clarify facts and intent before alleging misconduct by others. Recognizes that differences of opinion and honest error do not constitute unethical behavior.
2. Avoids condoning or appearing to condone statistical, scientific, or professional misconduct. Encourages other practitioners to avoid misconduct or the appearance of misconduct.
3. Does not make allegations that are poorly founded, or intended to intimidate. Recognizes such allegations as potential ethics violations.
4. Lodges complaints of misconduct discreetly and to the relevant institutional body. Does not act on allegations of misconduct without appropriate institutional referral, including those allegations originating from social media accounts or email listservs.
5. Insists upon a transparent and fair process to adjudicate claims of misconduct. Maintains confidentiality when participating in an investigation. Discloses the investigation results honestly to appropriate parties and stakeholders once they are available.
6. Refuses to publicly question or discredit the reputation of a person based on a specific accusation of misconduct while due process continues to unfold.
7. Following an investigation of misconduct, supports the efforts of all parties involved to resume their careers in as normal a manner as possible, consistent with the outcome of the investigation.
8. Avoids, and acts to discourage, retaliation against or damage to the employability of those who responsibly call attention to possible misconduct.

APPENDIX
Responsibilities of organizations/institutions

Whenever organizations and institutions design the collection of, summarize, process, analyze, interpret, or present, data; or develop and/or deploy models or algorithms, they have responsibilities to use statistical practice in ways that are consistent with these Guidelines, as well as promote ethical statistical practice.

Organizations and institutions engage in, and promote, ethical statistical practice by:

1. Expecting and encouraging all employees and vendors who conduct statistical practice to adhere to these Guidelines. Promoting a workplace where the ethical practitioner may apply the Guidelines without being intimidated or coerced. Protecting statistical practitioners who comply with these Guidelines.
2. Engaging competent personnel to conduct statistical practice, and promote a productive work environment.
3. Promoting the professional development and maintenance of proficiency for employed statistical practitioners.
4. Supporting statistical practice that is objective and transparent. Not allowing organizational objectives or expectations to encourage unethical statistical practice by its employees.
5. Recognizing that the inclusion of statistical practitioners as authors, or acknowledgement of their contributions to projects or publications, requires their explicit permission because it may imply endorsement of the work.
6. Avoiding statistical practices that exploit vulnerable populations or create or perpetuate discrimination or unjust outcomes. Considering both scientific validity and impact on societal and human well-being that results from the organization's statistical practice.
7. Using professional qualifications and contributions as the basis for decisions regarding statistical practitioners' hiring, firing, promotion, work assignments, publications and presentations, candidacy for offices and awards, funding or approval of research, and other professional matters.

Those in leadership, supervisory, or managerial positions who oversee statistical practitioners promote ethical statistical practice by following Principle G and:

8. Recognizing that it is contrary to these Guidelines to report or follow only those results that conform to expectations without explicitly acknowledging competing findings and the basis for choices regarding which results to report, use, and/or cite.
9. Recognizing that the results of valid statistical studies cannot be guaranteed to conform to the expectations or desires of those commissioning the study or employing/supervising the statistical practitioner(s).
10. Objectively, accurately, and efficiently communicating a team's or practitioners' statistical work throughout the organization.
11. In cases where ethical issues are raised, representing them fairly within the organization's leadership team.
12. Managing resources and organizational strategy to direct teams of statistical practitioners along the most productive lines in light of the ethical standards contained in these Guidelines.