

# Appendix 1: Principles Development Tool

This tool can be used for board discussions on ethics principles, as a working document for ethics boards developing an ethics code for AI, and as a tool to aid familiarity with AI ethics issues.

It provides a list of ethics principles that can be adopted by companies and other organizations. These are drawn from existing codes of AI and data ethics principles, internationally recognized human rights and labour rights documents, professional conduct codes for technologists, and statements of ethics principles by major technology companies.

These principles are suggestions for consideration. We recommend that organizations adopt a form of the eight main principles, which are either required under law or conventional human rights standards, are fundamental for the trust-building needed for AI success or both.

## How to use this tool

**Review the principles.** Consider whether they should be adopted as is, or with rewording. We suggest that principles be adopted if they are considered requirements (necessary under law, international human rights standards or contractual obligation) or recommendations (of benefit to your organization's business because they are consistent with the organization's values, increase trust by stakeholders, or follow the professional codes of conduct expected of employees and management) or go "above and beyond" your organization's self-interests to support social benefits that your organization favours.

**Consider adding additional principles.** Ask these questions:

- What are the most profound ethical issues facing the organization? Do the principles in this tool address them? If not, work backwards from those issues to identify the principles that should guide the people who face them.
- What are the long-term strategic goals of the organization pertaining to innovative technology? How will earning the trust of the organization's stakeholders support them? Are the principles in this tool sufficient to establish that trust? If not, work backwards from these goals to identify those principles.

**Consider the people impacted.** Ask these questions:

- Who are the people responsible for following the code? How will the code apply to them? Think of developers, operators and users of AI systems.

Also consider overseers such as board members, management and ownership:

- How should the code be adapted to your industry? Consider the broader network of vendors, partners and customers who will be affected, and those in professional roles, for example, doctors and nurses in the medical field, or judges, prosecutors and police in law enforcement. How should the code be written so they can apply it to their work?
- Who is vulnerable if your organization acts irresponsibly or unethically? (Not just customers and patients, but children, minorities, people with disabilities and others.) What guiding principles will protect them?
- How should the code account for different cultures and laws?

**Identify potential conflicts between ethical principles.** Ask these questions:

- In what circumstances could there be conflicts between principles, such as between legal requirements and human rights?
- What are ways to resolve the conflict? Resolutions may include qualifying or amending a principle, giving primary to one principle or choosing not to use AI in these situations.
- Does this conflict require further study or discussion?

**Review periodically.** Consider what new AI ethical issues have been encountered by your company. Also consider issues encountered by other organizations that might be applicable to yours.

Principles of AI ethics									
	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>SAFETY: Deliberate or inadvertent harm caused by AIs should be prohibited, prevented and stopped</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>			
1. AIs and people who work with them should preserve the physical safety and security of all who could be affected by AI									
2. AIs and people who work with them should protect physical and intellectual property that could be affected by AI									
3. People with AI responsibilities – lawmakers and law enforcers, and owners, developers, operators and users of AIs – should prohibit, prevent and stop the use of AI for criminal purposes									
4. People with AI responsibilities should prevent neglect that could cause AIs to do harm									
5. Owners and developers of AI algorithms should restrict open access to algorithms when misuse endangers health and safety									
6. Armed service members should prevent AIs from autonomously making decisions and taking actions that endanger non-combatants									
7. The use of employees or contractors to review emotionally repellent images and text should be avoided. If necessary, these employees and contractors should receive safety precautions, compensation and healthcare as would others performing high-risk work									
8. AIs should not be used to punish people or as an instrument of torture									

## Principles of AI ethics

	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>PRIVACY: AIs and people with AI responsibilities should protect personal and client data. Those who gather or share data with AIs or from AIs should seek and respect the preferences of those whom the data is about, including their preference to control the data</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>			
1. People and client organizations should be able to access, manage and control data held about them by a third party									
2. Before using data, AI systems should seek the informed consent of those who own the data and those whom the data is about. If consent is not possible, data should be anonymized									
3. At the point of data collection, organizations should disclose how they obtain personal data, what they use it for, how they may repurpose the data, with whom they share it, and the methods with which they analyse it									
4. Client and personal data should not be shared with third parties without their permission									
5. Data provenance: Those working with data should know its source and the trustworthiness of that source									
6. The privacy and security safeguards used to protect data should match the expectations of the subjects of the data									
7. Client and personal data should not be obtained or used in ways that violate rights and ethical principles									
8. Those protecting data should use effective cybersecurity methods and practices, and migrate to more effective ones as they are developed or become known									
9. AI should not be used to undermine privacy rights									

## Principles of AI ethics

	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>PRIVACY: AIs and people with AI responsibilities should protect personal and client data. Those who gather or share data with AIs or from AIs should seek and respect the preferences of those whom the data is about, including their preference to control the data</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>			
10. Those who build and operate AIs should be knowledgeable and vigilant about how data could be combined with other data to violate privacy preferences and prevent the use of their data in those ways									
11. AI owners should use and collect the minimum personal information necessary for operating AIs									
12. AI owners should not collect data unless there is a clear, immediate and achievable purpose. They should not collect data just because they may find a use for it in the future									
13. Owners of AI systems should oppose the weakening of privacy protections, such as backdoors, or weakening or limiting commercial encryption technologies									
14. Employees should honour confidentiality agreements between their employers and others									

## Principles of AI ethics

	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>EQUALITY: AIs should make only fair decisions consistent with human rights</b>	<b>X</b>	<b>X</b>		<b>X</b>	<b>X</b>	<b>X</b>		<b>X</b>	
1. AI systems should not limit opportunities to work, own property, receive education or care, or access products, services and public spaces, due to biased algorithms and data									
2. People with AI responsibilities – lawmakers and law enforcers, and owners, developers, operators and users of AIs – should prevent bias in AI systems in regard to gender, ethnicity, physical features, sexual orientation or political and religious views									
3. People with AI responsibilities involving legal decisions should prevent bias based on income in addition to the above									
4. AI should not be used for surveillance to deliberately or inadvertently restrict the rights of the above categories of individuals									
5. AI should not be used in a way that deliberately or inadvertently restricts cultures and lifestyles, unless it is needed to protect public health									

## Principles of AI ethics

	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>WELL-BEING: AIs should be used to support prosperity, health, democratic civic processes, personal freedom, goodwill, environmental sustainability, and the protection of children, people with disabilities, displaced people and other vulnerable populations</b>	<b>X</b>	<b>X</b>		<b>X</b>	<b>X</b>		<b>X</b>	<b>X</b>	<b>X</b>
1. AIs should be used as a force for good									
2. AI should be used to benefit and improve individual lives									
3. AIs should be prevented from undermining goodwill towards others by increasing hatred and contempt									
4. AIs should be prevented from undermining democracy, self-determination and civic processes									
5. AIs should be used to bring the underprivileged into the economy and the underrepresented into civic processes									
6. Companies and nations that gain immense wealth from AI should share their wealth with the impoverished									
7. Those who create the benefits of AI should make them available to all									
8. AIs that resemble humans and animals should not be targets for behaviour that would be cruel if enacted on living beings									
9. Governments, civic organizations, companies and schools should help people flourish alongside AI									

## Principles of AI ethics

	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>COMPREHENSION: The reasons for any AI decisions and actions should be understood well enough for humans to control AIs for consistency with ethical principles, and to make human accountability possible</b>					<b>X</b>			<b>X</b>	
1. Explainability: AI should provide explanations regarding how it arrives at decisions and suggestions									
2. Traceability: Investigators should always be able to discover the root cause if an AI causes harm									
3. Truthfulness: Operators and developers of AI should provide true, honest explanations to users and people when requested									
4. Validation: Operators of AIs should be able to confirm the accuracy of AI calculations and outcomes									
5. Auditability: Third parties should be able to review and confirm the accuracy of AI calculations and outcomes									
6. Transparency: The decisions and actions of AI systems, and the reasons behind them, should be understandable by all									
7. Ethical fit: Those responsible for AIs should only select algorithms that can be comprehended well enough to be ethically fit for use									
8. Recordability: AI systems should be equipped with an “ethical black box” that records information about the ethical considerations, and which ensures transparency and accountability									
9. Those developing, operating or using AIs, especially those in positions of authority, should be educated on AI ethical issues and how to work with AIs consistent with ethical principles									
10. AI researchers should honestly and accurately educate the public about their work; AI providers should provide honest and accurate information about their products, services and the data they use. Both should educate the public and support discussion of the implications for society									
11. AI researchers, providers and regulators should participate in public dialogue on the ethical, social and other implications of AI									

## Principles of AI ethics

	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>ACCOUNTABILITY: The responsibility for an AI's decisions and actions should never be delegated to the AI. People should take responsibility for following ethical principles when working with AI and be held accountable when AIs break ethical principles and voluntary obligations</b>					<b>X</b>			<b>X</b>	
1. By law or voluntarily entered agreement, an organization or person should always be held responsible and accountable for an AI and what it does									
2. Employees at companies that own AIs should be responsible for adhering to ethical principles as they build, operate or use AIs									
3. Employees at companies that own AIs should be responsible for carrying out the obligations of their employers, unless they break ethical principles									
4. People should not let AIs make the final decision to end life									
5. People should always decide whether and how to delegate decisions to AIs									
6. Agency: Humans should retain the capacity and responsibility to override the decisions and actions generated by an AI									
7. AI developers and operators should not use prisoners or children to perform AI related work, nor should AI work be associated with any form of slavery									

## Principles of AI ethics

	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>REMEDATION: Those with AI responsibilities should seek to be educated by people affected by their AIs. Workers, customers and others affected should have fair means to seek assistance or redress should AI endanger their livelihood, reputation or physical well-being</b>		<b>X</b>						<b>X</b>	
1. Through dialogue and research, AI researchers, providers and regulators should seek to understand the views and interests of those who are or may be affected by the AIs they develop, operate or provide									
2. AI practices should be reviewed for compliance with ethical principles									
3. AI providers and researchers should accept any scrutiny, debate and limitations imposed by authorities									
4. Workers should be consulted on the implementation, development and deployment of AI systems that affect their work and livelihood									
5. Workers should be able to access, manage and provide corrections to the data generated about them by AI systems									
6. When AI systems are used in HR procedures, workers should be able to receive explanations about how the AI made any recommendations or decisions									
7. Corporations are responsible for assisting workers displaced by AI, through retraining programmes, job-change possibilities, lifelong learning and access to social security									
8. The process of acquiring and evaluating AI systems should include checking whether deployment is consistent with workers' rights as they appear in human rights laws, international labour conventions and collective bargaining agreements									
9. Organizations should establish ways for whistleblowers to safely report AI systems or AI work that violates ethical principles or the law									
10. AI systems should not be exempted from laws, regulations and agreements protecting and compensating people and companies for damages									

Principles of AI ethics									
	Required:			Recommended:			Above and beyond:		
	By law	By human rights	By contractual obligation	Promotes organizational values	Promotes stakeholder trust	Follows professional codes	Shares wealth	Benefits society	Sustains the environment
<b>PROFESSIONALISM: AI researchers, scientists and technicians should follow high scientific and professional standards</b>					X	X			
1. AI science research should follow the same review processes and ethical constraints as other scientific research, without compromise, evasion or subversion									
2. AI research should be rigorous and reliable enough to earn the trust of the scientific community									
3. AI scientists and engineers should strive to build a professional culture of cooperation, trust and openness together									

## AI ethics codes consulted

### International bodies

- EC-HEGAI, European Commission's High-Level Expert Group on AI, <https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai> (link as of 18/8/19).
- EGESNT, European Group of Ethics in Science and New Technologies, <https://ec.europa.eu/research/ege/index.cfm> (link as of 18/8/19).
- IBHR, International Bill of Human Rights, <https://www.ohchr.org/Documents/Publications/FactSheet2Rev.1en.pdf> (link as of 18/8/19).
- World Economic Forum, “5 Core Principles to Keep AI Ethical”, <https://www.weforum.org/agenda/2018/04/keep-calm-and-make-ai-ethical/> (link as of 18/8/19).

### International labour organizations

- International Labor Organization, [https://www.ilo.org/wcmsp5/groups/public/---ed\\_emp/documents/publication/wcms\\_142317.pdf](https://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_142317.pdf) (link as of 18/8/19).
- UNI Global Union/The Future World of Work, [http://www.thefutureworldofwork.org/media/35420/uni\\_ethical\\_ai.pdf](http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf) (link as of 18/8/19).

## **Professional associations and universities**

- Association of Computing Machinery, <https://cacm.acm.org/magazines/2017/9/220423-toward-algorithmic-transparency-and-accountability/fulltext> (link as of 18/8/19).
- EEE, “Ethically Aligned Design”, report 2, <https://icps.gwu.edu/sites/icps.gwu.edu/files/downloads/Havens%20Ethically%20Aligned%20Design.pdf> (link as of 18/8/19).

## **AI declarations**

- Future of Life Institute/Asilomar Declaration, <https://futureoflife.org/ai-principles/> (link as of 18/8/19).
- Montreal Declaration, <https://www.montrealdeclaration-responsibleai.com/the-declaration> (link as of 18/8/19).

## **AI partnerships**

- FAT/ML, Fairness, Accountability and Transparency in Machine Learning, <http://www.fatml.org/resources/principles-for-accountable-algorithms> (link as of 18/8/19).
- Information Accountability Foundation, <http://informationaccountability.org/wp-content/uploads/Artificial-Intelligence-Ethics-and-Enhanced-Data-Stewardship.pdf> (link as of 18/8/19).
- Partnership for AI, <https://www.partnershiponai.org/tenets/> (link as of 18/8/19).

## **Companies**

- Accenture, [https://www.accenture.com/t20160629T012639Z\\_\\_w\\_\\_/us-en/\\_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf](https://www.accenture.com/t20160629T012639Z__w__/us-en/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf) (link as of 18/8/19).
- Google, <https://ai.google/principles/> (link as of 18/8/19).
- IBM, <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf> (link as of 18/8/19).
- Microsoft, <https://www.microsoft.com/en-us/ai/our-approach-to-ai> (link as of 18/8/19).