

# Ethics

Empowering AI Leadership

# Contents

- 3 Introduction
- 6 Responsibilities
- 7 Oversight
- 8 Agenda
- 9 Resources
- 10 Endnotes
- 11 Appendix



# Introduction

The potential of artificial intelligence (AI) to cause harm as well as create a greater good has aroused great interest in the field of ethics and made it one of the most publicized areas, as well as an emerging topic of focus within a board's oversight responsibilities.

Any information technology can be used for good or ill. But AI requires special attention because it increases the hazards of IT by enlarging both its scope and the scale of its impact. Moreover, with AI it is not just a matter of use once a product is deployed, as the use of an AI system is tightly linked to its design. Setting ethics standards for AI systems is necessary to avoid harm and increase well-being. Standards include the principles that AI systems and their developers, operators and managers should follow as they develop, operate and use AI systems and the data on which they depend. Standards also include the expectation of compliance with those principles by organizations and the law. AI ethics standards are necessary, but they are not enough: standards need to be taught, facilitated, absorbed, and possibly enforced in an organization's culture and workflow.

This module provides tools and information to help boards oversee the setting of ethics standards, and the establishment of an ethics board. For content on who and how boards oversee ethics and other AI decisions, please review the Governance module.

## AI's ethics hazards

AI presents new hazards because it has capabilities lacked by previous IT systems. Counterfeiters can now create realistic fake videos, and companies can build chatbots that are nearly indistinguishable from human interactions. Autonomous systems could make life and death decisions without human oversight. In general, AI systems are able to make or recommend decisions to human operators, some with very high stakes.

Because AI systems learn, they are vulnerable to being trained by deliberately or inadvertently biased data, and to developing and following decision-making models with hidden biases. For example, AI systems have been accused of rejecting female candidates for jobs and recommending disproportionate criminal sentences and policing of minority groups due to such biases.<sup>1</sup>

Many AI techniques resist explainability, rendering it difficult to pinpoint the reasons for a specific decision, and also to assess if the decision path crossed an ethical line. Certain kinds of machine learning-based AI systems can make decisions without human oversight, based on complex patterns beyond human comprehension, and thus make decisions that cannot be predicted. Humans may not be capable of overriding such AI systems when they make instantaneous decisions in real time, as with autonomous vehicles and aircraft.<sup>2</sup>

How human-machine interactions take place affects the risk management process. Accuracy improves as AI systems learn. There is a risk, however, that decisions made solely by technology are accepted by users without proper human oversight.

## Defining ethical imperatives

Technology companies, professional associations, government agencies, NGOs and academic groups have already developed many AI codes of ethics and professional conduct. While these can be helpful resources for organizations developing their own codes, they do not provide a universal solution. Organizations must develop their own.

While people may agree on broad concepts, the specifics of those concepts may differ. Ethics codes for AI systems may focus on transparency, privacy and benefitting society at large, but our review of existing codes shows that the definitions, or what is included under these terms, can vary.

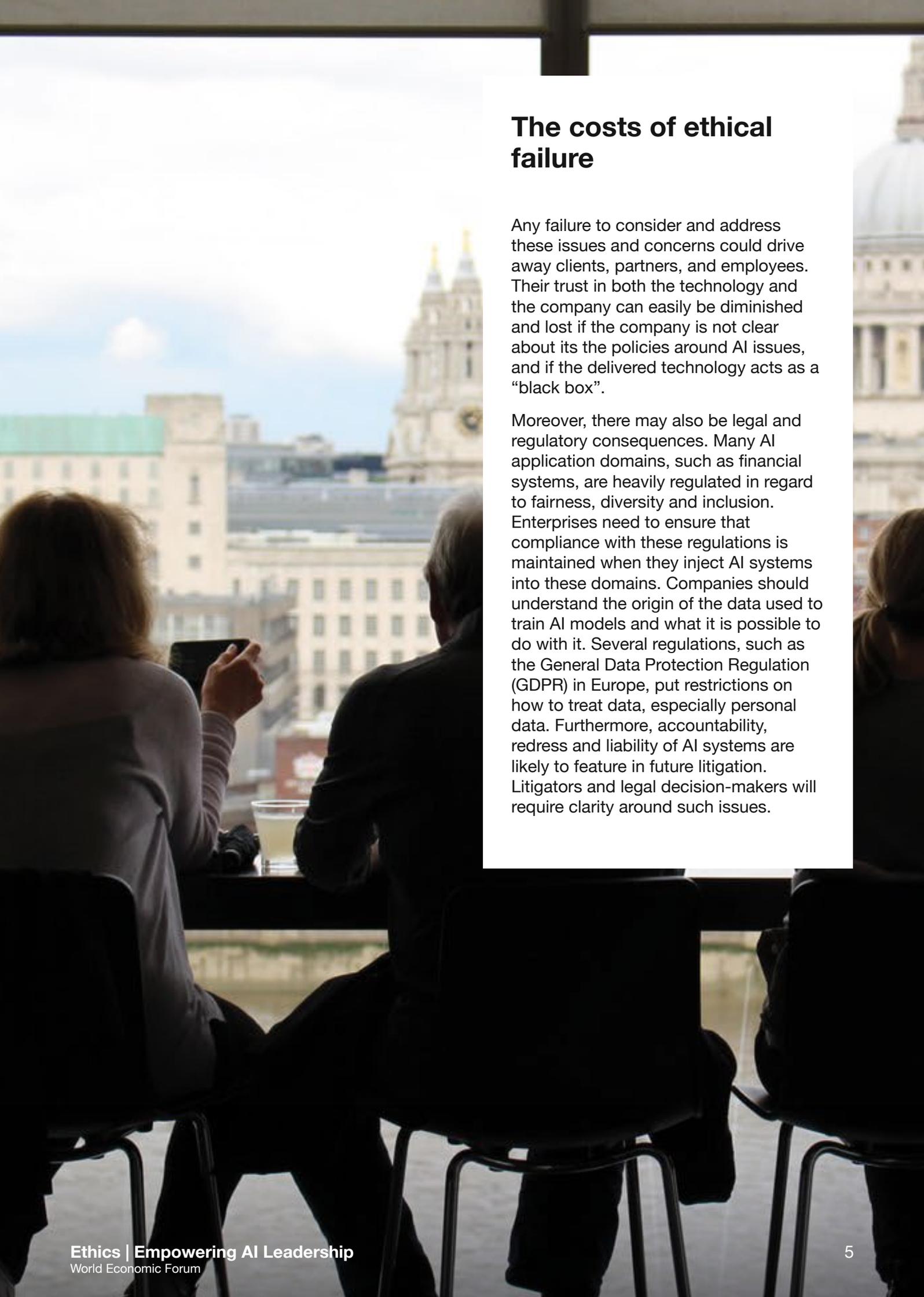
Public codes provided by technology companies are often aspirational rather than helpful guidelines. Some leave out specific commitments or define what a stated aspiration means. Academic codes are often aspirational in another way: They make specific commitments to public good, without spelling out how.

As organizations wrestle with developing codes, they will discover what moral philosophers already know: Ethical decision-making is often difficult. What is considered right and wrong can differ markedly by culture. People can view rights and legitimate interests from very different and sometimes conflicting perspectives, based on their experience and milieu. Some ethical dilemmas may be presented as a choice between lesser evils rather than of right from wrong. Ethics codes may be about making informed choices within guardrails that set limits of acceptable behaviour, and not simply about doing the right thing.

Codes address different audiences with different needs. Internally, codes must be the core of guidance mechanisms for making ethical decisions as organizations develop, deploy and use AI, such as codes of conduct and causes for dismissal. They must also be guarantors of employee rights and protections as well as their responsibilities. Externally, they serve as assurances to the public, as guarantees to customers and guidelines for vendors and partners. Effective AI ethics codes fulfil all these functions.

Defining the ethical principles that AI systems and their human creators and users should follow requires thoughtful analysis and consensus-building. So too does the implementation of those principles.

These issues must be thought through by leaders of each organization that deploys AI. While certain broad principles are found in nearly all codes, no one code is sufficient and appropriate for every organization.



## The costs of ethical failure

Any failure to consider and address these issues and concerns could drive away clients, partners, and employees. Their trust in both the technology and the company can easily be diminished and lost if the company is not clear about its the policies around AI issues, and if the delivered technology acts as a “black box”.

Moreover, there may also be legal and regulatory consequences. Many AI application domains, such as financial systems, are heavily regulated in regard to fairness, diversity and inclusion. Enterprises need to ensure that compliance with these regulations is maintained when they inject AI systems into these domains. Companies should understand the origin of the data used to train AI models and what it is possible to do with it. Several regulations, such as the General Data Protection Regulation (GDPR) in Europe, put restrictions on how to treat data, especially personal data. Furthermore, accountability, redress and liability of AI systems are likely to feature in future litigation. Litigators and legal decision-makers will require clarity around such issues.

# Responsibilities

Oversight of ethical standards for AI is a fundamental responsibility of the board.

**Our analysis of the G20/OECD Principles of Corporate Governance provides three main reasons:**

**The board ensures that ethics matter by hiring ethical executives and holding them accountable.**

According to Principle VI.C: “The board has a key role in setting the ethical tone of a company, not only by its own actions, but also in appointing and overseeing key executives and consequently the management in general. High ethical standards are in the long-term interests of the company as a means to make it credible and trustworthy, not only in day-to-day operations but also with respect to longer-term commitments.”<sup>3</sup>

**Boards cannot execute their responsibilities without ethical standards.**

Boards cannot effectively judge whether the company’s AI strategy, plans and performance are in alignment with their companies’ core values and ethical standards unless those standards are articulated. Without knowing what ethical standards should guide them or the dangers of violating them, boards cannot consistently “act on a fully informed basis, in good faith, with due diligence and care” (Principle VI.A). Nor can they “review corporate strategy, major plans of action, risk-management policies and procedures” for alignment to ethical standards, whether “setting performance objectives” includes meeting ethical standards, or whether ethical standards are met when “monitoring implementation and corporate performance, and overseeing major capital expenditures, acquisitions and divestitures” (Principle VI.D.1). Boards must therefore make sure these standards are set.

For more details about the ethical responsibilities pertaining to a specific strategy topic, please see the responsibility sections of the individual modules. For ethical responsibilities that are relevant across modules, please see the Responsibility module.

**Boards protect whistleblowers.**

Ethics standards are toothless if employees and other stakeholders cannot report violations to the highest authorities. “Stakeholders, including individual employees and their representative bodies, should be able to freely communicate their concerns about illegal or unethical practices to the board and to the competent public authorities and their rights should not be compromised for doing this” (Principle IV.E). To protect stakeholders’ ability to communicate ethics concerns regarding AI and data practices, the board should ensure the company establishes procedures and safe harbours for complaints by employees, their representative bodies and others outside the company, concerning illegal and unethical behaviour. The board should also give them direct and confidential access to an independent board member.

*The analysis in this section is based on general principles of corporate governance, including the G20/OECD Principles of Corporate Governance, 2015. It does not constitute legal advice and is not intended to address the specific legal requirements of any jurisdiction or regulatory regime. Boards are encouraged to consult with their legal advisers in determining how best to apply the principles discussed in this module to their company.*

# Oversight

## **This section includes five tools:**

The AI ethics principles development tool helps boards of directors and AI ethics boards develop an AI ethics code. It can also be used to guide board discussions on ethics principles. This tool contains eight broad principles, with more specific principles that follow on from them, and a way to evaluate their relevance.

*View the Principles Development tool in Appendix 1.*

Goals and guidance for the AI ethics board provides questions to consider before establishing an AI ethics board. It contains suggestions for goals for the AI ethics board to accomplish and guidance for it to follow. It also contains issues for the board of directors to consider in advance.

*View the AI Ethics Board Goals and Guidance tool in Appendix 2.*

Selecting the members of the AI ethics board suggests requirements to consider when appointing members to the AI ethics board. It contains four major requirements, along with questions to ask and actions to undertake during the search and evaluation process.

*View the AI Ethics Board Member Selection tool in Appendix 3.*

Assessing the draft AI ethics code provides questions to help directors evaluate the draft code presented by the AI ethics board.

*View the AI Ethics Code Assessment tool in Appendix 4.*

Assessing implementation, monitoring and enforcement of the AI ethics code includes questions to help boards evaluate whether they are receiving the information they require to carry out their oversight responsibilities, and whether the management team of the AI ethics board is effectively carrying out these responsibilities.

*View the Implementation, Monitoring and Enforcement tool in Appendix 5.*



# Agenda

The following suggestions can help individuals who prepare board discussions and set the board's agenda on ethics and AI:

## Before leading the first meeting

- **Prepare yourself:** Become familiar with the ethics issues created by AI, and the most urgent ethics issues for the board to address. The 'Resources' section provides reading on ethics issues. Speak to executives, in particular senior IT executive and security officers, about the ethics issues that concern them.
- **Gauge board member interest in AI risk:** Speak to other board members. Learn what importance they place on AI ethics and what concerns they have. Identify the board members who are most interested in achieving rapid progress on addressing AI ethics issues, and those who have concerns or lack interest.
- **Set goals:** Think ahead about the desired outcomes from the board discussion.

## Set the initial agenda items

These may comprise:

**Creating an AI ethics code.** Steps include:

- **Discuss:** Review the relevant AI ethics issues for your company, and your organization's readiness to address it (the Ethics principles development tool can help spark the discussion). The discussion should be informed not only by the organization's legal responsibilities and risks, but also by its ethical values and employee expectations.
- **Frame:** Discuss how the AI ethics code will be created: whether by a new or existing ethics board or committee, the resources it will be given, and how to assure its independence.

- **Set goals:** Consider the goals and guidelines for the board that will develop AI ethics guidelines.
- **Delegate:** Discuss who on the board will be responsible for establishing an AI ethics board, the qualifications of the AI ethics board members, and the goals and guidelines to set before them. These results should help meet the board's responsibilities, as recommended by the G20 OECD Principles.
- **Engage:** Discuss how the board will support the ethics code creation process without compromising the independence of the AI ethics board.

## Set follow-up agenda items.

These can include:

- **Review progress:** Discuss the work of the AI ethics board, including evaluating a draft AI ethics code, its implementation and its impact.
- **Access to information:** Discuss with fellow board members what information they require in order to stay current with emerging AI ethics issues.
- **Employee safeguards:** Discuss what steps to take to ensure the board is aware of illegal or unethical practices.
- **Review ethics programmes:** Periodically review the effectiveness of the ethics code and whether it needs to be updated.
- **Awareness of ethics issues:** Periodically review emerging and potentially significant ethics issues at your organization, the partners on whom the company relies and across your industry.
- **Broaden board thinking:** Invite an expert on AI ethics to present to the board. These could include academics or leaders from the organizations devoting themselves to AI ethics issues.

# Resources

(All references as of 18/8/19)

## Books

- Anastassia Lauterbach, *The Artificial Intelligence Imperative: A Practical Roadmap for Business*, Praeger, 2018.
- H. James Wilson and Paul R. Daugherty, *Human + Machine – Reimagining Work in the Age of AI*, Harvard Business School Press, 2018.
- Paula Boddington, *Towards a Code of Ethics for Artificial Intelligence*, Springer, 2017.

## Toolkits

- Markkula Center for Applied Ethics, Santa Clara University, “An Ethical Toolkit for Engineering/Design Practice”.

## Articles and reports

### AI ethics for boards

- Deloitte, “The Board’s Role in Ethics and Compliance”.
- Peter Collins, “We Need a More Open Debate on AI and Ethics in the Boardroom”, Centre for Ethical Leadership.
- Sabine Vollmer, “The Board’s Role in Promoting an Ethical Culture”, *Journal of Accountancy*.
- Trooper Sanders, “How a Strong Board of Directors Keeps AI Companies on an Ethical Path”, VentureBeat.

### AI ethics for boards

- Accenture, “Facilitating Ethical Decisions throughout the Data Supply Chain”.
- Accenture, “Is Explainability Enough? Why We Need Understandable AI”; “Responsible AI and Robotics, an Ethical Framework”; “The Responsible AI Imperative”.
- Annette Zimmermann and Bendert Zevenbergen, “AI Ethics: Seven Traps”, (link as of 18/8/19).
- European Economic and Social Committee, “Ethics of Big Data”.
- European Group on Ethics in Science and New Technologies, “IBE, Business Ethics and Big Data”.
- IEEE, “Ethically Aligned Design”.
- IBM, “Everyday Ethics for Artificial Intelligence: A Guide for Designers & Developers”.
- OECD, “Going Digital”.
- PwC, “Responsible AI and National AI Strategies”, for the EU Commission.
- World Economic Forum, “Top 9 Ethical Issues in AI”, 2016.

### Panels and boards

- Australian Computer Society’s AI Ethics Committee.
- Axon’s AI Ethics Board.
- DeepMind’s Ethics and Society fellows (AI research arm of Alphabet).
- Lucid’s Ethics Advisory Board.
- Microsoft’s AETHER Committee.

# Endnotes

*(All references as of 18/8/19)*

1. Karen Hao, “This is How AI Bias Really Happens – and Why It’s So Hard to Fix”, MIT Technology Review, 4 February 2019; Daniel Cossins, “Discriminating Algorithms: 5 Times AI Showed Prejudice”, New Scientist, 12 April 2018, updated 27 April 2018.
2. European Group on Ethics in Science and New Technologies, Statement on Artificial Intelligence, Robotics and “Autonomous” Systems, Brussels, 9 March 2018.
3. Karen Hao, “This is How AI Bias Really Happens – and Why It’s So Hard to fix”.

World Economic Forum  
91-93 route de la Capite  
CH-1223 Cologny/Geneva  
Switzerland  
Phone: +41 (0) 22 869 1212

San Francisco Centre for the Fourth Industrial Revolution  
1201 Ralston Ave, San Francisco, CA 94129,  
Phone: (415) 704-8848

[contact@weforum.org](mailto:contact@weforum.org) | [www.weforum.org](http://www.weforum.org)

For more information, contact Kay Firth-Butterfield,  
Head of AI and Machine Learning,  
Centre for the Fourth Industrial Revolution.

[kay.firth-butterfield@weforum.org](mailto:kay.firth-butterfield@weforum.org)

# 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).

# Appendix 2: Goals and Guidance for the AI Ethics Board

Before establishing an AI ethics committee to create a set of AI ethics principles, decide what the board will be expected to accomplish, and the guidance to give them. Preparing this in advance will help the board of directors recruit members for the ethics board, and select members who can achieve these goals. The following is a list of goals and guidance for the board of directors to consider when establishing the committee.

## Before beginning

Before establishing an ethics board, the board of directors should discuss the following:

- **AI ethics concerns:** What issues does the board of directors wish the AI ethics board to address? What needs – can be articulated to the AI ethics board members?
- **Readiness:** How ready is the organization to address ethical issues as they pertain to technology? Who is aware of AI ethics issues among the management team and employees? Have people in the organization begun to experience AI's ethical quandaries? How well does the company understand and address the ethics issues it has faced up to this point?
- **Existing boards:** Does the organization currently have an ethics board? If yes, how should the AI ethics board work with the existing board as it draws up a set of AI ethics principles?
- **Permanence:** Are AI ethics best served by a permanent body devoted to those issues, a permanent ethics body with responsibilities for other ethics issues, or a temporary task force? How will compliance with the AI ethics code be effectively monitored and enforced in the future?
- **Independence:** What steps should the board of directors and management take to ensure the AI ethics board maintains its independence? In addition to its selection of the board (see Appendix 3), how should the board of directors ensure that the AI ethics board members will stay independent?
- **Resources:** Will the AI ethics board require a budget and its own staff to carry out its work? How large a budget should be provided, and from what funds?

## Goals

- **Recommend whether to update existing codes, develop a new code specifically for AI, or a broader code covering technology and data.** This includes identifying whether codes now exist, evaluating their effectiveness, and whether AI deployment requires a code for the entire company or just a small number of individuals.
- **Develop an actionable set of principles.** While broad and simple principles are memorable, they should be specific enough to provide meaningful guidance for people making real and often difficult decisions. This includes translating broad principles to specific guidelines for different kinds of employees, managements and suppliers.

- **Recommend benchmarks and performance indicators.** This includes recommending which ethical AI design, development and research standards the organization should follow, key performance indicators of ethical behaviour, and which ethics benchmarks or rankings to measure the organization against.
- **Suggest how to help people learn and comply with the code.** Recommend ways to make the code easier and more convenient to understand and apply, such as ethics hotlines and advisers, ethics training and online ethics chatbots.
- **Recommend how frequently the code should be updated.** Suggest a minimum period that should pass before revisiting the code, and triggers that would require a code update before that period ends. Triggers could include new laws and regulations, the emergence of new AI technologies and capabilities, the development or purchase of new AI systems, new labour agreements and acquisitions of companies. The ethics board should also ask whether the organization is equipped to update standards, and what needs to be done to improve readiness.
- **Recommend whether to appoint an ethics officer specifically for AI.** Consider whether the company should establish such a role rather than appoint a chief AI ethics officer or add AI to the current ethics officer's portfolio. If so, recommend what their responsibilities would be and how to judge their success.
- **Recommend how to resolve potential conflicts between principles.** The ethics board could identify situations where principles may conflict with laws or other principles, who would be affected, and recommend a course of action for reconciling or choosing between the conflicting elements. These may include conflicts between employee and customer protections or between human rights standards and local laws.
- **Identify any parts of the code that should not be made public.** Consider whether there are parts of the code that would violate agreements with suppliers or customers if they were made public.
- **Recommend how the code should be implemented.** This includes who should review the code besides the board and management team, whether the code should be piloted before being finalized, how to educate people about the code, and creating a timeline for roll-out and implementation.
- **Suggest whistleblowing mechanisms:** This includes avenues by which employees, contractors, customers and providers can safely report ethics violations and problems.

## Guidance

- **Support the organization's principles and goals.** The recommendations of the AI ethics board should be consistent with the organization's principles and support the goals of the board of directors in terms of: organizational survival; achievement of the organization's mission; increasing the value created by the company; compliance with laws and regulations; meeting mandated obligations.
- **Support the greater good.** Consistent with the World Economic Forum's commitment to improving the state of the world, ethics boards should also consider including principles within the AI ethics code that go beyond the organization's self-interest and benefit humanity and the planet as well.

- **Respect human rights:** Consistent with the UN Guiding Principles on Business and Human Rights, the AI ethics code:
  - should help companies avoid causing or contributing to adverse human rights impacts through their own activities
  - should seek to prevent or mitigate such impacts when the enterprise is directly linked to them via a business relationship
  - should engage in ongoing due-diligence processes to identify, prevent and mitigate salient human rights risks
  - should provide remediation for human rights impacts through legitimate mechanisms when adverse human rights impacts do occur
- **Be consistent:** Ethics codes should be internally consistent. The principles of the code should follow the same logic, support rather than undermine one another and be applicable to all. The provisions of the ethics code should apply definitions equally, and guiding principles should support one another. The code should be applicable to all who are affiliated with the organization: no leaders, staff or AI users should be exempt. Ask whether different groups of employees or the public would find them to be inconsistent.
- **Meet external standards:** At a minimum, the codes should meet the ethical standards set forth in law, and of the most important customers. The code should also meet or exceed the standards of the company’s industry associations, its vendors and partners, and the professional associations of its employees – or have defensible reasons if it does not.
- **Be defensible:** Codes should be able to withstand public scrutiny. They should be written so that they are as transparent as possible to the public. Even sections that are judged as inappropriate for release (for instance, for financial reasons) ought to be defensible if they were made known.
- **Be clear and easy to use:** The code should be clearly written without requiring technical knowledge. It should also be organized so that any reader can easily find sections pertinent to their situation. The committee should also consider how to make the code more convenient to implement.
- **Protect the vulnerable:** Children can succumb to addictive behaviour when algorithms learn how to gain and retain their attention; unlike adults, they are not able to learn why and thus change their behaviour. Minorities can lose out on economic opportunities or be treated unfairly due to biases in algorithms and datasets. Ethics codes should protect these and other vulnerable populations.
- **Seek outside voices:** The ethics committee should seek out the views of people who are or will be affected by the organization’s use of AI. This may require inviting outsiders to address meetings, holding open hearings where the public can address the ethics committee and soliciting written statements.
- **Consider local conditions:** Codes must be operable wherever the organization conducts business. While difficult to achieve, ethics guidelines for AI should be flexible enough to accommodate local cultures and laws without compromising core principles. Ethics boards should be prepared to suggest how to find the best balance for their organization.
- **Consider industry requirements:** Similarly, codes must be operable within the company’s business. They should be written so they are plainly applicable to work processes and the specific kinds of decisions made in the company. Ethics codes must be especially careful to distinguish discriminatory and non-discriminatory activity – for example, AI decisions that legitimately deny services to a member of a group from decisions that discriminate due to deliberate or inadvertent bias against that group.”

# Appendix 3: Selecting the Members of the AI Ethics Board

The AI ethics board should consist of members who meet four main requirements. While selecting the board members, consider the following:

**Independent:** The ethics code will be far more credible if the AI ethics board members are not beholden to personal interests that affect their judgement. When creating the board and choosing ethics board members, ask whether ethics board members:

- Are free to make their own recommendations on ethics principles
- Are not under pressure to adopt specific recommendations from board members
- Are free of outside financial or family interests that might influence their judgement or give the appearance of conflicts of interest

**Individually knowledgeable, collectively representative:** Ethics boards require a mix of expertise and empathy. They should give voice to the people who have to work under these standards, and for those who could suffer the consequences if AI is managed irresponsibly. While creating the AI ethics board, ask whether the mix of members includes:

- People who are knowledgeable about the company's existing ethics codes and codes of conduct. These might include compliance officers and human resource managers.
- Experts in ethical decision-making. These might include professional ethics consultants and academic experts on ethics.
- Experts in AI who are knowledgeable about the ethical issues that have and will arise in the field: e.g., AI experts and computer scientists from academia or consulting firms who have studied AI ethics issues.
- Experts in the ethical issues faced by management teams and employees in the company wherever the organization operates, and those faced by its partners and competitors. These might include compliance officers, financial executives and function or business process experts inside or outside the organization.
- Legal experts who know the legal and regulatory requirements in the jurisdictions where the company operates.
- People who can ably represent those affected by AI decisions and actions such as customers, employees or patients; children and others unable to represent themselves; women, minorities and others who could be affected by bias and other AI ethics problems. These representatives should include labour union representatives and employees; they could also include members of professional associations and advocacy organizations. There should be opportunities for input and comment on drafts, as being undertaken by other AI ethics development bodies.

**Of good moral standing:** Ethics boards should not include people of questionable integrity. When selecting individuals to serve on the board, the same level of review when hiring selecting senior managers is required:

- Conduct a reference check
- Check for criminal records
- Review the quality of their work on, or relating to, AI ethics issues

**Of constructive character:** Members of ethics boards should have a personal character that is appropriate for an ethics board. While performing reference checks, enquire whether the candidate:

- Is a good listener who respects and accommodates other points of view
- Can apply ethical principles to new and complex situations
- Has a reputation for treating other people respectfully

# Appendix 4: Assessing the Draft AI Ethics Code

After the AI ethics board delivers its draft AI ethics code and recommendations, evaluate the results against the goals set at the beginning of the process. In addition, consider asking these questions when approving the code:

- Did the AI ethics board follow the guidance provided by the board of directors?
- Does the AI ethics code address the reasons an ethics code is needed?
- Does it recommend changes to the organization's older codes of conduct?
- Is the code flexible enough to keep pace with changes in AI and the business, but specific enough to be useful to the people who will apply it?
- Is the AI ethics code consistent with internationally recognized human rights principles, as enshrined in the UN Guiding Principles on Business and Human Rights and the International Bill of Human Rights?
- Is the AI ethics code of conduct directly linked to the company's ethical values, beyond legal and regulatory requirements?
- Have the company's reputation and trust, employee engagement, inclusion and diversity goals been upheld by the AI ethics code?
- Are there mechanisms for dealing with new situations not explicitly addressed in the AI ethics code? Or for addressing conflicts between the code's principles?
- How are individuals motivated or reminded to comply with ethical AI standards?
- Are there conflicts between the principles in the proposed code of ethics and those of partners and customers? Does it contain the ability to resolve these conflicts?
- Does the AI ethics code use external sources on AI ethics and perspectives?

# Appendix 5: Assessing Implementation, Monitoring and Enforcement of the AI Ethics Code

Ask the following questions after implementation of the AI code of ethics is underway or complete. These questions can be used whether it is an AI ethics board, an ethics board with broader scope or management that is responsible for implementation, monitoring or enforcement.

**Implementation of the code:** Does the board of directors receive (from the management team or ethics board):

- Insight on how the business units are implementing the AI ethics code?
- Reactions from employees and management to the implementation of the code?
- Plans of action for addressing implementation issues?

Are there processes in place (by the management team or the ethics board) to:

- Allow all stakeholders to express their views on the implementation of the code?

**Monitoring adherence, compliance, responses and new developments affecting AI ethics:** Does the board of directors receive responses to the following questions (from the management team or AI ethics board):

- Is there access to relevant, timely and complete information about the company's AI ethics activities?
- Are there regular updates on compliance with the code? Do these updates include measures, such as performance against KPIs or ethics benchmarks?
- Are there continuous and, when necessary, real-time updates on emerging AI ethics issues that have been raised by clients, partners and competitors, and from within the company?

Are there processes in place (by the management team or ethics board) to:

- Measure and report compliance with the code?
- Determine how clients and partners evaluate the organization's AI systems for ethics?
- Monitor acquisitions and partnerships for adherence to the AI code of ethics?
- Allow employees, customers and other stakeholders to express their opinion about ethical issues with the AI systems deployed by the company? Or about the impact of AI on existing systems?
- Check whether any of the conditions that trigger an update of the code have occurred?

**Enforcing the code:** Has the management team:

- Included AI ethics activities within the annual business plan?
- Set aside sufficient funds and other resources to support monitoring and enforcement?

Does the board of directors:

- Include compliance to the AI ethics code when evaluating executive performance, compensation and termination?
- Review acquisitions, sell-outs and other business actions for their impact on the organization's AI ethics values and adherence to the ethics code?