

Appendix 1: Examples of Technology Ethics and Responsible Technology Principles

As noted previously, there are many examples of principles in use at technology companies, as well as sources for principles from academia and elsewhere. Here are just a few sets of corporate ethics principles, as well as some ways that they are implemented at those organizations.

Microsoft's Responsible AI Principles

To promote better outcomes for new technologies and reinforce social trust, Microsoft has developed six AI ethics principles, to guide their thinking and judgment:

Fairness: AI systems should treat all people fairly

Reliability & Safety: AI systems should perform reliably and safely

Privacy & Security: AI systems should be secure and respect privacy

Inclusiveness: AI systems should empower everyone and engage people

Transparency: AI systems should be understandable

Accountability: People should be accountable for AI systems⁶¹

Microsoft also has several associated tools for thinking about ethics during product development:

Judgment Call, the Envision AI Workshops, Impact Assessment, Community Jury, and a set of various tools for ethically improving machine learning models. More details on Microsoft's work on responsible AI can be found in [Responsible Use of Technology: The Microsoft Case Study](#).⁶²

IBM's Principles and Pillars

For IBM, their principles are “the guiding values that distinguish IBM's approach to AI ethics.” There are 3 of these principles:

⁶¹ Microsoft, “Our Approach,” *Microsoft website*, 2022, available at: <https://www.microsoft.com/en-us/ai/our-approach?activetab=pivot1%3aprimar5>

⁶² WEF and Markkula Center for Applied Ethics at Santa Clara University, “Responsible Use of Technology: The Microsoft Case Study,” *The World Economic Forum website*, February 2021, available at: <https://www.weforum.org/whitepapers/responsible-use-of-technology-the-microsoft-case-study>

The purpose of AI is to augment human intelligence: At IBM, we believe AI should make all of us better at our jobs, and that the benefits of the AI era should touch the many, not just the elite few.

Data and insights belong to their creator: IBM clients' data is their data, and their insights are their insights. We believe that government data policies should be fair and equitable and prioritize openness.

Technology must be transparent and explainable: Companies must be clear about who trains their AI systems, what data was used in training and, most importantly, what went into their algorithms' recommendations.⁶³

IBM also has five Pillars that are “our foundational properties for AI ethics.” These 5 pillars are:

Explainability: Good design does not sacrifice transparency in creating a seamless experience.

Fairness: Properly calibrated, AI can assist humans in making fairer choices.

Robustness: As systems are employed to make crucial decisions, AI must be secure and robust.

Transparency: Transparency reinforces trust, and the best way to promote transparency is through disclosure.

Privacy: AI systems must prioritize and safeguard consumers' privacy and data rights.⁶⁴

IBM operationalizes their principles and pillars through several “toolkits,” which directly help AI developers make their models more explainable, fair, robust, and so on: the AI Explainability 360 toolkit, the AI Fairness 360 toolkit, the Adversarial Robustness 360 toolkit, the AI FactSheets 360 toolkit, the AI Privacy 360 toolkit, the Uncertainty Quantification 360 toolkit, and the Causal Inference 360 toolkit.⁶⁵ More details can be found in [Responsible Use of Technology: The IBM Case Study](#).⁶⁶

Salesforce's Core Values and Ethical Use Guiding Principles

Salesforce prides itself on being a company driven by values since its inception. It has core values and ethical use guiding principles. Its core values are:

⁶³ IBM, “AI Ethics,” *IBM website*, 2022, available at: <https://www.ibm.com/artificial-intelligence/ethics>

⁶⁴ Ibid.

⁶⁵ IBM, “Trusted AI,” *IBM website*, 2022, available at: <https://research.ibm.com/teams/trusted-ai>

⁶⁶ WEF and Markkula Center for Applied Ethics at Santa Clara University, “Responsible Use of Technology: The IBM Case Study,” *The World Economic Forum website*, September 2021, available at: https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_The_IBM_Case_Study_2021.pdf

Trust: We act as trusted advisors. We earn the trust of our customers, employees, and extended family through transparency, security, compliance, privacy, and performance. And we deliver the industry's most trusted infrastructure.

Customer Success: When our customers succeed, we succeed. So we champion them to achieve extraordinary things. We innovate and expand our business offerings to provide all our stakeholders with new avenues to achieve ever greater success.

Innovation: We innovate together. Our customers' input helps us develop products that best serve their business needs. Providing continual technology releases and new initiatives gives our customers a competitive advantage.

Equality: Everyone deserves equal opportunities. We believe everyone should be seen, heard, valued, and empowered to succeed. Hearing diverse perspectives fuels innovation, deepens connections between people, and makes us a better company.

Sustainability: We lead boldly to address the climate emergency. We are committed to bringing the full power of Salesforce to accelerate the world's journey to net zero.

Salesforce's ethical use guiding principles are:

Human Rights: We work to ensure the direct use of our technologies upholds equal and inalienable protections.

Privacy: We push the frontier of privacy best practice in our product design to enable customers to protect individuals' data.

Safety: We aim to protect humans from direct harm from the use of our technology.

Honesty: We oppose the use of our technology to knowingly spread disinformation or conspiracy theories.

Inclusion: We create opportunity through equal access to technology.⁶⁷

In order to coordinate and implement their values and guiding principles, Salesforce created an Office of Ethical and Human Use of Technology. Salesforce also incorporated these values and principles into their overall goals through their corporate strategy and accountability mechanism, the "V2MOM" – for vision, values, methods, obstacles and measures. More details on Salesforce's work, including their product-specific guiding principles, can be found in [Responsible Use of Technology: The Salesforce Case Study](#).⁶⁸

⁶⁷ Salesforce, "Ethical Use Policy," *Salesforce website*, 2022, available at:

<https://www.salesforce.com/company/intentional-innovation/ethical-use-policy/>

⁶⁸ WEF and Markkula Center for Applied Ethics at Santa Clara University, "Responsible Use of Technology: The Salesforce Case Study," *The World Economic Forum website*, September 2022, available at:

https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_Salesforce_Case_Study_2022.pdf

Google's Principles

Google has seven AI Principles:

- 1. Be socially beneficial.**
- 2. Avoid creating or reinforcing unfair bias.**
- 3. Be built and tested for safety.**
- 4. Be accountable to people.**
- 5. Incorporate privacy design principles.**
- 6. Uphold high standards of scientific excellence.**
- 7. Be made available for uses that accord with these principles.**

Google also has identified four AI applications it will not pursue:

- Technologies that cause or are likely to cause overall harm. Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks and will incorporate appropriate safety constraints.
- Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.
- Technologies that gather or use information for surveillance violating internationally accepted norms.
- Technologies whose purpose contravenes widely accepted principles of international law and human rights.

At the end of their page, Google leaves open the chance to revise their principles as they gain more experience, saying: “As our experience in this space deepens, this list may evolve.”⁶⁹

For implementation, Google recommends certain practices for AI, saying that “Reliable, effective, user-centered AI systems should be designed following general best practices for software systems, together with practices that address considerations unique to machine learning. Our top recommendations are outlined below, with additional resources for further reading.”

Google's recommended practices include:

- Use a human-centered design approach
- Identify multiple metrics to assess training and monitoring
- When possible, directly examine your raw data
- Understand the limitations of your dataset and model
- Test, Test, Test
- Continue to monitor and update the system after deployment⁷⁰

⁶⁹ Google AI, “Artificial Intelligence at Google: Our Principles,” *Google AI website*, 2022, available at: <https://ai.google/principles/>

⁷⁰ Google AI, “Responsible AI practices,” *Google AI website*, 2022, available at: <https://ai.google/responsibilities/responsible-ai-practices/>

[Google has used Markkula Center resources](#) to shape their thinking when it comes to ethical AI, particularly the [Ethics in Technology Practice](#) materials.⁷¹

In summary, each of these companies has developed both principles to guide their choices and tools to help operationalize ethical thinking. Principles without operationalization are just nice words: the proof of ethical intentions is in the actions and outcomes themselves.

⁷¹ Kent Walker, “Google AI Principles updates, six months in,” *Google Blog*, Dec 18, 2018, available at: <https://www.blog.google/technology/ai/google-ai-principles-updates-six-months/> and Shannon Vallor, Brian Green, and Irina Raicu, “Ethics in Technology Practice,” *Markkula Center website*, June 22, 2018, available at: <https://www.scu.edu/ethics-in-technology-practice/>