A practical guide to implementing AI ethics governance

Capgemini

Foreword

With the rapid acceleration in the pace of AI adoption, it has never been more vital for organizations to ensure that they are executing their transformation based upon clear ethical guidelines. AI has the potential to transform our world in incredible ways, but if implemented without a framework, and the associated risk management processes, it has an equal chance of causing significant, even irreparable, harm. At Capgemini, we have had the framework that is outlined in the follow pages in place for many years, with a robust and proven Code of Ethics for AI at its heart.

Every organization's code of ethics principles should be unique to them, reflecting their values and broader societal and cultural norms. With that in mind, the AI Futures Lab team have created this thorough and practical guide to help you develop or validate the right code of ethics principles for AI for your organization, as well as guidance on how to establish the broader risk management and governance processes.

These steps will help you to move forward with your business transformation with confidence, ensuring that together we are delivering the future we all want.

Anne-Violaine Monnié
Group AI Ethics Leader, Capgemini

Where to start?

Ethical conversations can be complicated. They depend on cultural, societal, personal, and organizational values. There's no such thing as a one-size-fits-all set of **ethics*** principles for organizations. While other reports may try to provide you with a readymade set of generic **AI ethics*** principles, this guide will instead provide you with a toolkit to create a set of AI ethics principles that are unique to your organization. It will also provide you with key insights on the responsibilities and processes that will be needed to ensure they have the desired impact.

Ethics is not a concept born in the 21st century. For centuries, people have explored and debated moral frameworks to guide their lives, societies, and innovations. With that perspective, the following pages present a timeline highlighting key moments in this ongoing journey.

Monika Byrtek
AI Philosopher, AI Futures Lab, Capgemini

James Wilson
AI Ethicist, AI Futures Lab, Capgemini

***Ethics -** (1) a set of moral principles : a theory or system of moral values (2) the principles of conduct governing an individual or a group (Webster Dictionary)

***AI Ethics -** (1) a set of moral principles related to the use of AI (2) the field that applies moral principles and values to guide use of AI systems

Table of contents

A necessity, not a luxury

The increasing need to adopt enterprise-scale AI within organizations has amplified ethical concerns. Amongst these concerns are dilemmas that could have far-reaching consequences for organizations, customers, and even society as a whole. This practical guide intends to provide a perspective on the foundational steps required to implement an effective governance framework for the ethical challenges organizations may encounter during and after their AI transformation.

While headlines often dramatize AI risks, the real threats are often subtle – like biased credit algorithms that use gender to determine credit limits. Without early intervention, these ethical risks can rapidly turn into serious legal and reputational consequences. As generative and agentic AI enter mainstream usage, the need for organizations to embed agile, ethical governance to mitigate evolving challenges becomes more apparent. These innovations add substantial complexity in terms of maintaining adequate explainability and control. This need only further amplifies when you consider emerging risks from innovations like embodied AI and quantum computing.

The shifting AI landscape

Before the launch of ChatGPT in November 2022, enterprise-scale AI focused on narrow, well-controlled use cases that were overseen by qualified individuals and teams who could validate its outputs. For example, computer vision-based diagnostic X-Ray tools would be monitored by qualified radiographers. Though ethical concerns existed, the impact of AI was relatively non-intrusive to most people's lives.

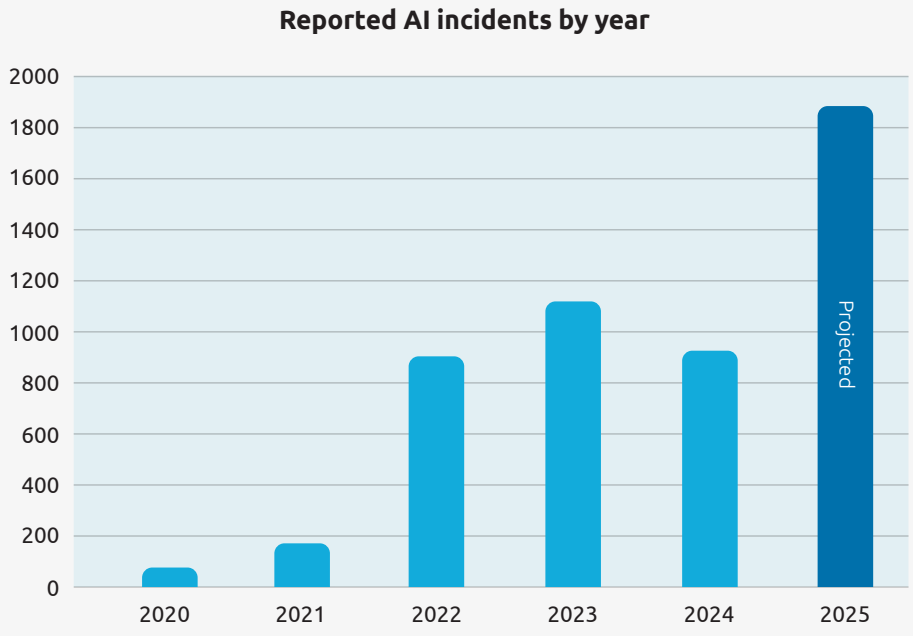


Figure 1: Source the AI Incidents Database

Thought Experiment: Imagine a story about AI causing harm across society appears on the front page of your local newspaper or news site. How would you feel if your company was implicated in that article? Now turn this scenario around and also ask yourself, how would you feel if you were the victim of unethical AI-augmented actions?

A timeline of ethics in humanity

Hammurabi: Born in 18th century BC

The Code of Hammurabi is one of the earliest intact sets of ethical/legal principles

Socrates: Born ca. 470 BC

The concept that all virtues are from knowledge

Everything changed with the launch of ChatGPT. AI became widely accessible, and frontier model providers and hyperscalers began releasing increasingly powerful tools which required complex oversight. Early missteps were mostly harmless (e.g. chatbots selling cars for $1 or giving advice on how to glue cheese to pizza), but as capabilities grew, so did the risks. In 2024, a companion chatbot was directly implicated in at least one teenage suicide, while another encouraged a failed assassination attempt. The risk profile for future AI adoption is only compounded further by embodied AI and action models, which are AI models that can interpret and act upon real-world physics concepts.

Regulations are evolving around the world and this alone isn't enough to fully protect society from AI harm. **Legality doesn't necessarily guarantee ethical integrity.** For example, consider the algorithm used by a government entity to allocate child-care payments. The algorithm made its decisions based on proxy data related to ethnic characteristics such as surname. **Conversely, there are instances where activities are considered illegal but still ethical,** for instance, gray-hat hacking, or whistleblowing[1] on unscrupulous activity by your employer may be highly ethical but can also be considered illegal if it transgresses the terms of a non-disclosure agreement you have in place with them.

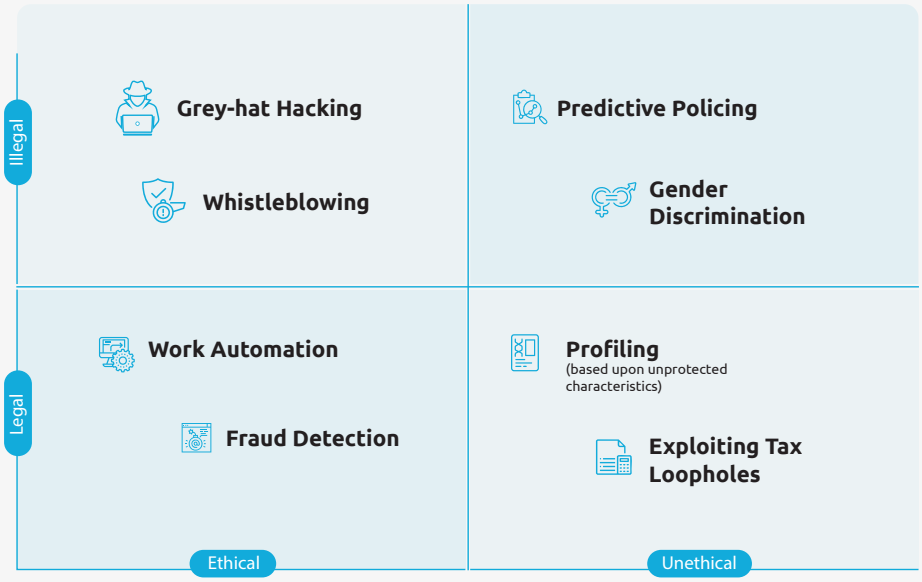
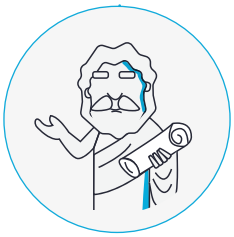


Figure 2: Examples of the intersection between legality and ethics

It is therefore critical that organizations apply equal rigor to the ethical, legal, and robustness aspects of their AI governance. While legal counsel is a standard part of transformation **programs, organizational ethics reflect what a company believes is right – not just what is required by law.**

Ideally, organizational values would align with those of the cultures and societies in which the organization operates, but alignment across global operations is complex. Not all employees share the same ethical lens, as this is shaped by lived experiences and cultural norms. This reinforces the need for a clear, leadership-driven value framework to define and guide employee behavior at work.



Plato: Born ca. 430 BC

Created the famous "Allegory of the cave"

Aristotle: Born ca. 384 BC

Believed that people should achieve excellent character by practicing virtue

Epicurus (Epicureanism): Born 341 BC

Considered the absence of pain as a greatest pleasure that people should strive for

1. It should be noted that Whistleblowing is legal in many countries.

However, when adopting AI platforms with third-party components, organizations are effectively onboarding "virtual employees"[2] with no lived experience or intrinsic ethics. Their behavior is shaped entirely by pre-training on undisclosed data and is constrained only by external guardrails. As such, there's no guarantee that their actions will reflect an organization's values[3]. This absence of alignment to expected cultural, societal, and organizational values means that we must provide them externally to the system via internal guardrails and governance.

Evolving geopolitical dynamics are also shaping AI development, which remains concentrated in the United States, China, and Europe – regions that don't fully reflect global cultural diversity. Models trained in one region may yield outputs only relevant to that specific cultural context. While this may sometimes be correct, in other cases it can lead to significant ethical issues. To ensure relevance and effectiveness, we need a wide variety of cultures to be represented in large language models (LLMs). A key challenge to this is the limited access to diverse linguistic and cultural resources for training.

These factors have led to the rise of the **AI ethicist role**, which is already commonplace within many organizations undergoing AI-focused business transformation.

Zeno of Citium (Stoicism): Born 334 BC

Lived by the concept of stoic calm

Cyrenaic school (Hedonism): Existed ca. 4th century BC

Taught that pleasure is the most important principle in life

Thought Experiment: What should the role of an AI ethicist be when someone in your organization proposes the implementation of an AI solution with clear ethical risks. For instance, using a third-party black box AI to determine customer credit limits, where the provider refuses to disclose the training process or data sources? What support should your organization's ethics principles for AI provide in such a case?

2. Important Author's note: Intentional anthropomorphism: We have intentionally inserted a few examples of anthropomorphism within this guide to help visualise some key points around AI behaviour. Where this is the case, we have put the text in quotes e.g., "virtual employee".

3. In reality there is no guarantee that the actions of real employees will be fully aligned to your organizational values, and in fact well-managed AI could be effectively more "obedient" than any human.

What is an AI ethicist and why do you need them?

It is important to first clarify what an AI ethicist isn't:

AI ethicists are not the moral arbiters for the organization. Ultimate responsibility for AI ethics must sit with the executive team, just as with any other business outcomes. However, the AI ethicist plays a key role in asking the right questions around AI implementation and clarifying risk ownership. As articulated by Olivia Gambelin, a pioneer in Responsible AI, *"Ethicists must possess the ability to neutrally observe an ethically charged situation, abstract the details of such out to the higher ethical principles at play in order determine right from wrong, and then bring this decision back down to contextually specific actions."**

AI Ethics is a team sport. The AI ethicist must ensure that diverse and inclusive perspectives are considered before decisions are made. This includes cultural, educational, and experiential diversity. For example, Capgemini's AI Futures Lab's ethics team includes a classically trained philosopher as well as technologists and business experts. Ethics governance is a team sport. At a minimum, having a well-rounded team will help to counteract individual decision-making biases. As Professor Toby Walsh's research shows, at least 21 common cognitive biases can influence decisions if left unchecked.

At least 21 forms of human bias (and counting)

1. Anchoring	8. Hindsight Bias	15. Systemic Bias
2. Belief Bias	9. Information Bias	16. Risk Compensation
3. Confirmation Bias	10. Loss Aversion	17. Selection Bias
4. Distinction Bias	11. Normalcy Bias	18. Time-saving Bias
5. Endowment Effect	12. Omission Bias	19. Unit Bias
6. The Framing Effect	13. Present Bias	20. Zero-Sum Bias
7. The Gambler's Fallacy	14. The Recency Illusion	21. Flatpack Bias

Figure 3: From "Machines Behaving Badly: The Morality of AI"[4] - Toby Walsh - ISBN: 978-0750999366

Mahayana Buddhism: Emerged ca. 1st century BC

There are 5 principles of moral life: dana (charity), virya (fortitude) sila (morality), ksanti (patience), and dhyana (meditation)

Augustine of Hippo: Born 354 AD

First christian attempt at compiling moral philosophy

Thomas Aquinas: Born 1225 AD

Cardinal virtues: prudence, temperance, justice, and fortitude and theologicalvirtues: faith, hope, and charity

* Gambelin, O. Brave: what it means to be an AI Ethicist. AI Ethics 1, 87–91 (2021).

4. The Flatpack Bias involves favouring things built by yourself

From principles to practice

Ethical AI is essential for every organization. Even if you aren't deploying AI directly, your partners, suppliers, or customers are likely to be doing so. Like data protection, AI ethics must be embedded across the organization. Assuming executive buy-in, the next step is to define an operating model for AI ethics that functions at every level of the organization. At its core is the AI ethicist – whose responsibilities may vary, but will always include ensuring ethical risks are addressed, ownership is clear, and diverse perspectives are considered.

Building AI ethics capabilities – core responsibilities within the AI ethicist's role:

1. **Define and maintain a code of ethics principles for AI** aligned with organizational values, applied across all AI initiatives.
2. **Collaborate with governance teams** to track alignment with the organization's code of ethics principles for AI and associated targets (e.g., fairness metrics). Maintain an AI inventory, and conduct regular risk and impact assessments, and monitor the AI landscape in your company.
3. **Stay current on AI developments,** including evolving risk profiles and global incident reporting.
4. **Work with the leadership team to embed AI ethics governance** into transformation programs, including the interfaces across legal, security, data, and delivery teams.
5. **Establish an ethical AI operating model,** including guidance on engaging broader communities regarding ethical dilemmas and maintaining risk profiles.
6. **Conduct AI ethics training** as part of broader AI literacy efforts across the workforce.
7. **Identify and drive valuable accreditation programs** e.g., ISO-42001
8. **Drive consistent messaging** on the value of ethical AI, highlighting reputational and business benefits.
9. **Create safe channels for raising employee concerns** about AI use and ethical risks.
10. **Promote AI ethics principles** internally and externally to boost your organization's profile and reputation.

The first step is to work with delivery and governance teams to define AI ethics principles that are practical and aligned to organizational values. The next section explores key concepts that should be addressed.

Thought Experiment: A new project is in the pipeline that would increase annual revenue by $1bn, but the AI ethicist has warned that it could also lead to a 20% reduction in the workforce and relies on a new open-source dataset that your organization does not control. What actions should your organization consider to ensure that you stay aligned to your ethical principles before making a decision about whether to proceed?

Thomas Hobbes: Born 1588 AD

Beginnings of social contract theory

René Descartes: Born 1596 AD

I think therefore I am – "Cogito, Ergo Sum"

Establishing the right AI ethics principles

Ethics principles for AI provide a mechanism for managing and assigning ownership of AI risk within the organization. According to a Gartner report from October 2024, the five most common principles are: **human-centric and socially beneficial, fair, explainable and transparent, secure and safe, accountable.**

You may refer to Capgemini's own Code of Ethics for AI, however it's essential that each organization defines its own principles – rooted in its own values and unique context. A strong starting point is performing a SWOT analysis of the key risks and challenges that AI adoption may pose to your organization, or society more broadly. This should span four dimensions: technological, psychological, sociological, geopolitical. The analysis not only informs the initial principle design but also serves as a foundational artifact for ongoing governance. Though the SWOT analysis may begin as a design tool, it should remain a living part of your AI governance framework. As part of the governance process, the following questions should be asked at a minimum:

- How will we build on the strengths?
- How will we address our weaknesses?
- How will we capitalize on the opportunities?
- How will we avoid the threats?
- How will we use our strengths to mitigate our weaknesses?
- How will we exploit the opportunities to offset the threats?
- What are our priorities?

Ethical AI governance process at a glance

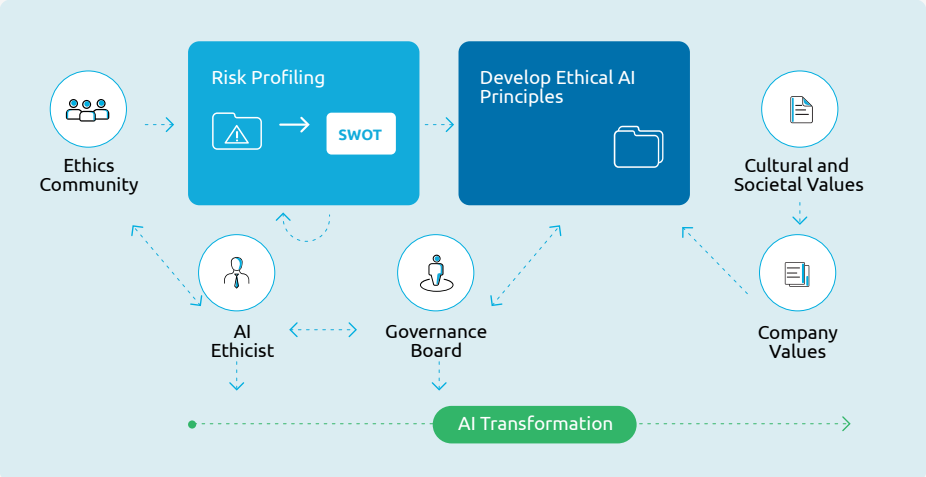


Figure 4

Thought Experiment: Are any of the following principles included in your code of ethics for AI? Human-centric and socially beneficial, fair, explainable and transparent, secure and safe, accountable.

Baruch Spinoza: Born 1632 AD

Substance, attributes and modes. Further exploration of the concept of rationalism

John Locke: Born 1632 AD

Government exists by the consent of the people to protect natural laws: life, liberty, and property

Jean-Jacques Rousseau: Born 1712 AD

Naturalism and the concept of the noble savage

The next section contains a **sample SWOT analysis**, along with a narrative describing the rationale for why some of these risks have been included.

A holistic framework for managing ethical AI risks

AI impacts and risks-a sample SWOT analysis

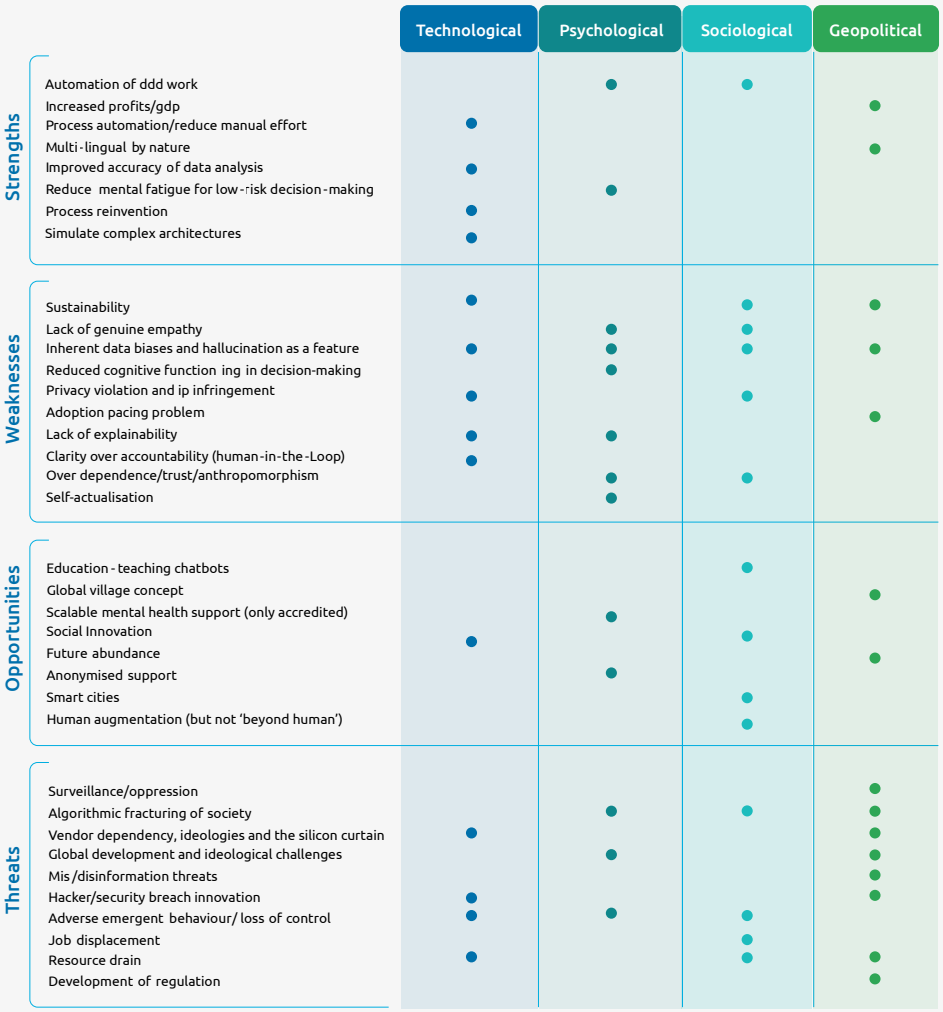


Figure 5: A sample SWOT analysis

Thought Experiment: What impact(s) do you want your organization's use of AI to have on the organization itself, the workforce, the local community, and the environment?

Immanuel Kant: Born 1724 AD

The categorical imperative encourages to act in such a way that it could, and also will, become a universal law

Georg Wilhelm Friedrich Hegel: Born 1770 AD

Social institution to provide concrete content to moral duties

Technological concepts explained

Ethics principles for AI must address several technological concepts, many of which underpin the psychological, sociological, and geopolitical concepts that will be further discussed on the next pages.

Data and processing

Beyond legal compliance, the principles define what data is appropriate for training and operating AI systems. Balance the benefits of broader datasets (e.g., improved accuracy) against privacy risks. Principles should also clarify the organization's stance on using models that have been trained with data for which the vendor may not have appropriate intellectual property rights.

Accuracy and hallucination

While narrowly trained AI models can sometimes outperform humans in specific tasks, generative AI is prone to erroneous responses (including hallucinations) and communicates these very convincingly. Ethics principles for AI must ensure these risks are mitigated through design, delivery, and ongoing monitoring.

Transparency/ explainability

Given the varying degrees of explainability available through different models, all AI decisions must be justifiable, transparent, and readily interpretable by those they impact. Data inputs and outputs, business rules, and user interactions must be addressed as part of this process.

Managed bias

Bias is one of AI's most complex ethical challenges (see the section *Considering bias and fairness*). Systems must be designed to detect and address unwanted bias.

Environmental impacts

AI, especially Gen AI, consumes significant energy and water. Principles should ensure sustainable design, efficient use of infrastructure, and active monitoring of environmental impact. Supply chains must also be evaluated to ensure ethical and sustainable sourcing of labor and materials.

Adverse emergent behavior in autonomous agentic AI

As multi-agent autonomous systems become more complex, they may exhibit emergent behaviors – some beneficial and others unpredictable or misaligned with ethical standards. The ethics principles must account for collective behaviors that may not be evident when monitoring individual agents.

Accountability and human-in-the-loop

AI systems need to have clear human accountability. While human-in-the-loop is standard for critical decisioning made using AI, agentic AI – capable of autonomous action – requires enhanced oversight. This includes increased auditability, transparency, and human-OVER-the-loop controls. Accountability is non-negotiable, and those responsible must own the impacts of the AI's actions, including any legal or reputational implications. Additionally, organizations must maintain rollback capabilities and retain critical human skills to manage system failures.

Arthur Schopenhauer: Born 1788 AD

Morality stems from compassion

John Stuart Mill: Born 1806 AD

Creator of the scientific method and supporter of utilitarianism

Karl Marx: Born 1818 AD

Beginning of the concept of materialism and materialistic ethics

Psychological concepts explained

AI adoption can affect individuals psychologically. These factors should also inform your ethics principles:

Trust, overdependency, and anthropomorphism

Generative AI can "lie" in a very convincing and engaging manner (see point on *Accuracy and hallucination*), and users are naturally inclined to anthropomorphize technology. This combination can lead to misplaced trust and overreliance. Ethics principles must require clear AI identification and safeguards against overdependence.

Mental wellbeing

Building on the human-centric design concept, consider how AI processes impact users' cognitive load (e.g., increased information flow or faster business "tickspeed" can heighten stress and pressure). In addition, consider any potential adverse impacts on the workforce during development of the solution, for instance in any outsourced activity to create or moderate training data. Exploitation of low-cost labor in developing regions is a documented issue and must be addressed ethically.

Critical thinking and decision making

AI should support – not replace – human judgment. Users must be provided with the evidence, access, and time needed to apply critical thinking and ensure confident, informed decision-making.

Friedrich Nietzsche: Born 1844 AD

The superman (Übermensch) and the will to power

Isaac Asimov: Born 1920 AD

Three laws of robotics written in 1942

Sociological concepts explained

An organization's success heavily depends on its societal impact – from customer trust to employee well-being. Ethical AI must reflect this context. Key areas to consider include:

Education and literacy

Ensure users, customers, leaders, and investors understand how to confidently and effectively use AI systems. AI literacy and critical thinking skills are essential to informed use and decision-making.

Maintaining privacy

Design AI systems that uphold strict data privacy standards, protecting both user and customer data, as well as any sensitive or regulated information.

Fairness

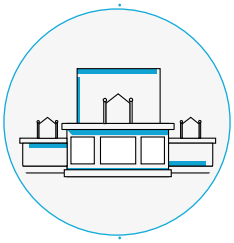
AI must operate fairly across all user groups. This includes involving diverse and inclusive teams in design and testing to identify and manage unwanted bias.

Human-centric design

Consider how AI impacts roles – especially when automating human job functions. One benefit of AI adoption is the removal of human labor that is dull, dirty, or dangerous (DDD). While automation can improve safety and efficiency, organizations must consider how it reshapes responsibilities and creates new opportunities. Failure to support this transition risks losing critical skills and diminishing job satisfaction.

Loss of control

Whether through a lack of operational control over AI solutions or the more recent suggestions that Large Language Models have exhibited self-preservation and deceit in an effort to disobey or ignore their human controllers, there is a growing risk that organizations could lose control of their AI implementations, with the potential for catastrophic outcomes.



Nuremberg Code: 1947 AD

One of the most important documents related to clinical research ethics

Declaration of Human Rights: 1948 AD

Adopted by UN, relates to rights and freedoms of all human beings

Alan Turing: Born 1912 AD

Writes 'Computing Machinery and Intelligence' 1950

Geopolitical concepts explained

While ethical AI must account for individual and societal impacts, it also plays a critical role in protecting organizational operations and reputation within a global context. Some key focus areas include:

Compliance with laws and regulations

The principles must ensure AI systems comply with local laws across all operating regions. Regulations like data protection, privacy, and sovereignty laws reflect societal values, which vary significantly between the EU, U.S., and China. For example, consider how your AI implementation might unintentionally enhance a state's surveillance capabilities of groups or individuals.

Silicon curtain and vendor dependency

As the geopolitical landscape changes, access to AI technologies may become restricted – a risk that historian and philosopher Yuval Noah Harari calls "the drawing of the Silicon Curtain." Organizations must assess how their reliance on specific vendors, models, or infrastructure could be affected by such a shift.

Global development

Consider how your ethics principles for AI can promote equality, improve individual and community welfare, and avoid deepening the wealth divide.

Ideological challenges

AI models may reflect the biases of their developers and the political and economic systems within which they're built. While no model is entirely ideology-free, organizations must evaluate the impact of these influences in their implementations.

Algorithmic fracturing and feedback loops

Like filter bubbles or echo chambers on social media, AI systems can reinforce narrow worldviews. It's important to consider how this phenomenon might impact your organization, particularly given the risks highlighted in the *"Ideological challenges"* section above.

Thought Experiment: What would you like the legacy of your organization to be in your local community or even wider? Which values would support this?

Declaration of Helsinki: 1964 AD

Ethical principles regarding medical research involving human participants

Joseph Weizenbaum: Born 1923 AD

ELIZA – The first Chatbot created ca. 1966

Considering bias and fairness

Bias and fairness sit at the heart of any ethical AI framework – and they're among the most complex challenges to address. While it is often assumed that bias is harmful and should be eliminated, that's not always the case.

Destructive bias example

An algorithm that reviews and filters job applicants for a role based upon training on a dataset of existing or historical employee profiles that is itself biased based upon protected characteristics such as age range, gender, or ethnicity.

Necessary bias example

A model predicting the likelihood of Sickle Cell Anaemia (SCA) deliberately overrepresents individuals of African descent, as they are disproportionately affected. In this case, the bias reflects a real-world medical reality that is essential for accuracy.

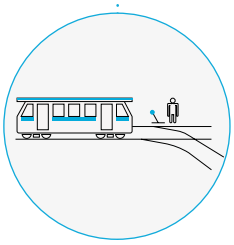
Rather than removing all bias, organizations should focus on fully understanding the biases and ensuring that the associated risks are owned by the accountable individuals within the organization.

The butterfly effect?

There is a common, but erroneous assumption that technology should be neutral and using a computational process like AI to automate decision-making should allow you to eliminate unwanted bias and create a completely fair system. In practice, however, this assumption is flawed. The very act of identifying and correcting known biases raises the risk of unintentionally accentuating other unclassified biases within the decisioning process.

This "can't see the forest for the trees" scenario can compromise the critical thinking required for system design. Focusing on this sub-set of identified biases can initiate a domino effect, where attempting to address one bias can make others worse or potentially introduce new, emergent ones. For example, focusing solely on resolving an age range bias for a committee may inadvertently introduce a previously unconsidered bias around gender. It is also worth considering Simpson's Paradox, which illustrates how trends visible in groups of data points can disappear, or even reverse, when multiple groups are aggregated. As an example, the data may show a distinct preference for different colors of T-shirts, based upon a gender-based analysis, however this distinction disappears when the datasets are aggregated. This may be caused by disproportionate representation of demographics within the individual groups.

These complexities highlight the need for caution, diverse perspectives, and continuous critical thinking when addressing bias. Managing bias isn't a one-time fix – it's an ongoing process, especially as AI systems evolve and interact in new ways.



Philippa Foot: Born: 1920

Creator of the trolley problem ethical thought experiment in 1967

Arthur C. Clarke: Born 1917 and Stanley Kubrik: Born 1928

Introduces us to HAL in '2001 AD: A Space Odyssey' in 1968

Belmont Report: 1978 AD

Ethical principles and guidelines for the protection of human subjects in research

Testing the principles

Using the outputs of the SWOT Analysis alongside company values, the AI ethicist, in collaboration with key functions (e.g., Legal, Enterprise Architecture, and HR) and selected representatives from the broader ethics community, can start shaping ethics principles tailored to the organization. These principles should be tested across the full AI lifecycle and evaluated against emerging technologies that may impact the business (as per the example below). Consider for instance, whether the principles address the introduction of AI agents acting as "virtual employees" with their own autonomy, authority, and agency? Are they prepared for embodied AI (AI-enabled robotics) or the future implications of quantum computing?

Given the rapid evolution of the AI landscape, it's unrealistic to expect principles to be perfect or fully future-proofed from the outset. Instead, treat them as living guidelines that are designed to evolve. This process is intended to be repeated on a regular basis, perhaps annually, to ensure that your principles remain relevant and practical.

Table 1

SDLC phase	Ideation	Sourcing	Development	Testing	Utilisation	Retirement
Principle 1	😐	😐	😐	😐	😐	😟
Principle 2	😐	😐	😐	😐	😐	😐
Principle 3	😐	😐	😐	😐	😐	😐
Principle 4	😐	😟	😐	😐	😐	😐
Principle 5	😐	😐	😐	😐	😟	😐
...	...	...	...	...	...	...
Principle n	😐	😐	😐	😟	😐	😐

Table 2

Innovation	Agentic AI	Robotics	...	Quantum
Principle 1	😟	😐	...	😐
Principle 2	😐	😐	...	😐
Principle 3	😐	😐	...	😐
Principle 4	😐	😟	...	😐
Principle 5	😐	😐	...	😐
...	...	...	...	...
Principle n	😐	😐	...	😐

Figure 6: Testing the principles against the project lifecycle and innovation pipeline

Next steps

We are operating in a rapidly evolving AI landscape. New opportunities and risks emerge daily, requiring organizations to be agile while strategically planning for the future. AI solutions must be legal, ethical, and robust. Clear AI ethics principles are no longer optional, they are essential for delivering innovative solutions **that create lasting value for both business and society.**

Creating an ethical environment means embedding these principles across all levels of the business – from executive leadership to teams developing client-facing solutions and internal processes. This alignment is critical in the age of AI.

In this guide, we emphasized the importance of establishing **AI principles that are unique to the organization's individual context, culture, and history.** The principles should address both current risks and opportunities as well as those on the horizon. This process is a team sport, which should reflect voices and ideas from all levels of the organization as well as the communities impacted by the adoption of AI. Most importantly, it requires active support from senior leadership to ensure the development of **purposeful, valuable, and ethical principles, processes, and tools.**

Jean Baudrillard: Born 1929 AD
Writes about simulation theory in 1981

James Cameron: Born 1954 AD
Movie: 'The Terminator' released in 1984

OpenAI unleashes ChatPT on the World: 2022 AD.

?

Recommended reading

The AI ethicist role is still relatively new in many organizations, however there are a number of reports and books that can help shape the thinking and practice for the role. Below are a few suggestions curated by the authors of this report:

- Confidence in AI Playbook
- Confidence in Autonomous and Agentic Systems
- What is an AI Ethicist, and Why do we need them?
- Capgemini AI Labs "Ethics and More" Newsletters
- Artificial Negligence by James Wilson
- The AI Strategy Manifesto by Bora Ger
- Technology is Not Neutral by Dr. Stephanie Hare
- Machines Behaving Badly: The Morality of AI by Toby Walsh
- Human Rights, Robot Wrongs by Dr. Susie Alegre
- AI Ethics by Mark Coeckelbergh
- Human Compatible by Prof. Stuart Russell
- Privacy is Power by Carissa Veliz
- Brave: what it means to be an AI Ethicist by Olivia Gambelin
- Nexus by Yuval Noah Harari
- Weapons of Math Destruction by Cathy O'Neil
- Unmasking AI by Joy Buolamwini

AI Futures Lab

We are the AI Futures Lab - expert partners that help you confidently visualize and pursue a better, sustainable, and trusted AI-enabled future. We do this by understanding, preemptying, and harnessing emerging trends and technologies. Ultimately, making possible trustworthy and reliable AI that triggers your imagination, enhances your productivity, and increases your efficiency. We will support you with the business challenges you know about and the emerging ones you will need to know to succeed in the future.

Build your AI advantage, layer by layer. Backed by extensive research and collaboration, we're best placed to help you navigate the AI landscape, and establish AI solutions that herald a step change in how we can solve business problems, holistically. Engage with us – let us surprise you with our visionary mix of what's to come.

Contacts

Robert Engel

Head of AI Futures Lab
robert.engels@capgemini.com

Mark Roberts

Deputy Head of AI Futures Lab
mark.roberts@capgemini.com

Bora Ger
bora.ger@capgemini.com

Jonathan Aston
jonathan.kirk@capgemini.com

Monika Byrtek
monika.byrtek@capgemini.com

Niharika Kalvagunta
niharika.kalvagunta@capgemini.com

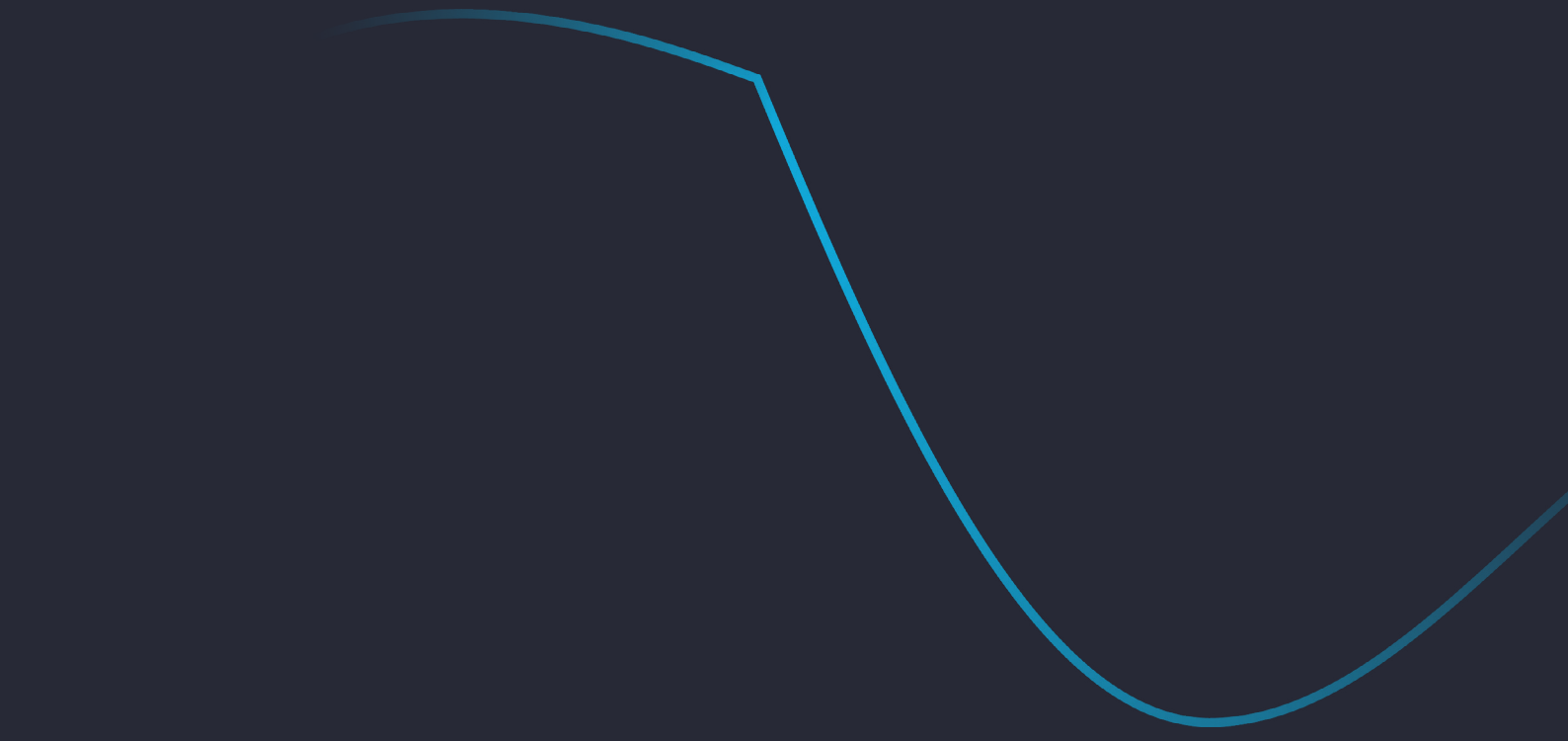
Johan Müllern-Aspegren
johan.mullern-aspegren@capgemini.com

James Wilson
james.l.wilson@capgemini.com

About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 350,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fuelled by its market leading capabilities in AI, generative AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2024 global revenues of €22.1 billion.

Get the future you want | www.capgemini.com



Capgemini