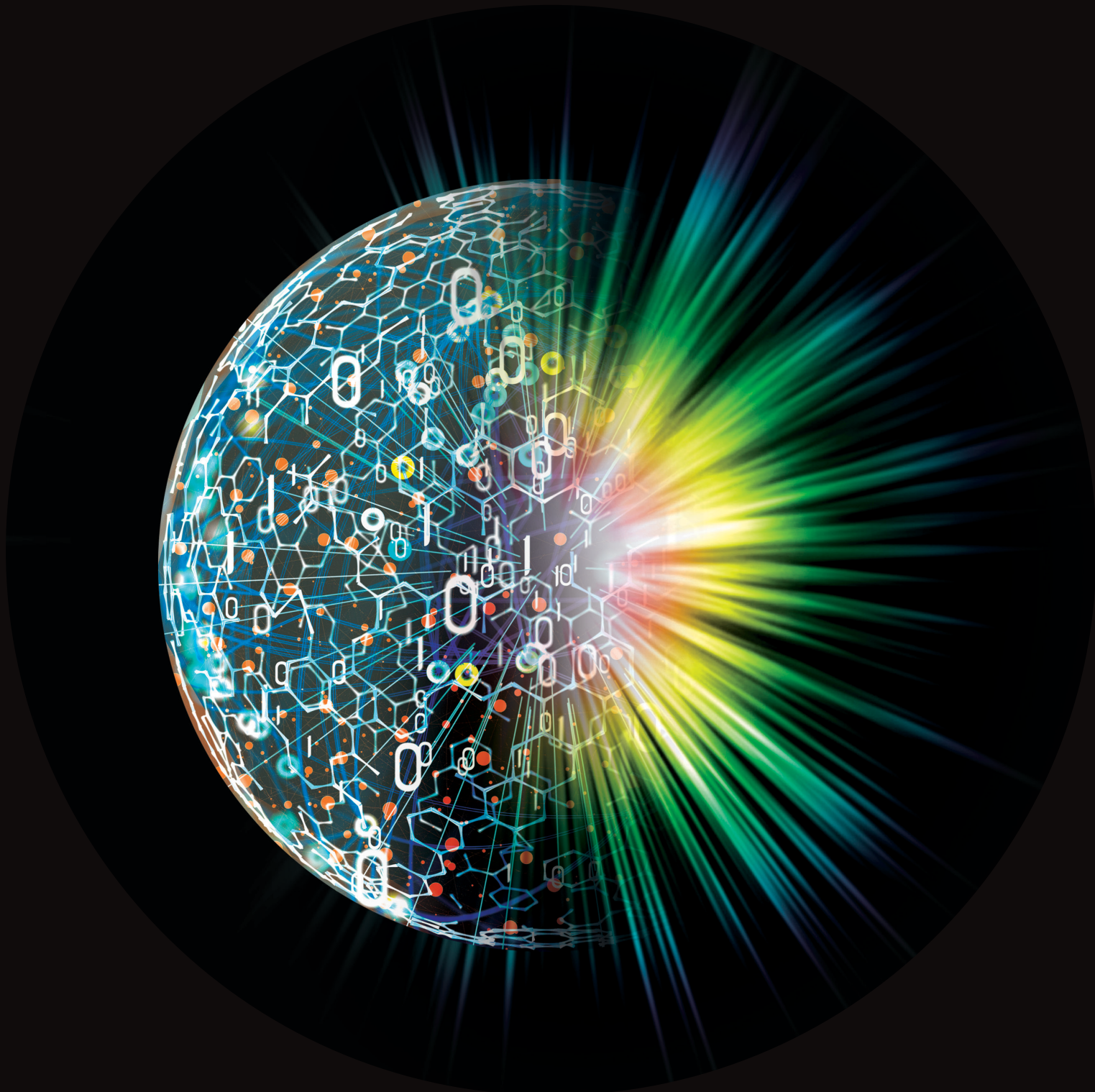


**Deloitte.**



**AI ethics**

A new imperative for businesses,  
boards, and C-suites



# Introduction

For individuals, leaders, organizations, and society, an increasingly urgent question is whether artificial intelligence (AI) can help us advance in positive ways, or will it hinder us?

The application of AI as a problem-solving tool offers great promise for advancement. AI in medicine, for instance, could **throw skyrocketing**

**medical costs in reverse**. AI in agriculture could lead to **double or triple the crop yields** of today's inputs, a pivotal achievement in the ongoing fight against food insecurity.

But AI can also be used against our human interests through such means as cyberattacks, **social manipulation or discrimination**, and **competing financial incentives**. Realizing this, consumers and citizens are concerned about the accumulation of data being used to make decisions without consumer consent or respect for consumer rights. New laws like the European

Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act have been promulgated, at least in part, to address these concerns.

---

Ethical risks are a chief concern for about a third of executives who expect near-term business transformation from AI.<sup>1</sup>

---

Near the end of his life, physicist Stephen Hawking—who famously spoke of AI's benefits and dangers—coauthored an **open letter** urging AI researchers and

developers to focus their work in an ethical direction. Not long afterward, 76 percent of executives responding to a **Deloitte survey** said they expected AI to “substantially transform” their companies within three years. About a third of this group added that ethical risks were a chief concern about AI technology.

But what does an ethical direction in AI look like, and how can organizations put it into action? That is the subject of the discussion to follow. Let's begin with some basic definitions.

# AI ethics, defined

AI is a broad term encompassing technologies that can mimic intelligent human behavior.<sup>2</sup> Four major categories of AI are in increasingly wide use today:

- **Machine learning**, the ability of statistical models to develop capabilities and improve their performance over time without the need to follow explicitly programmed instructions.
- **Deep learning**, a complex form of machine learning used for image and speech recognition and involving neural networks with many layers of abstract variables.
- **Natural language processing (NLP)**, technology that powers voice-based interfaces for virtual assistants and chatbots, as well as querying data sets, by extracting or generating meaning and intent from text in a readable, stylistically neutral, and grammatically correct form.
- **Computer vision**, technology that extracts meaning and intent out of visual elements, whether characters (in the case of document digitization) or the categorization of content in images such as faces, objects, scenes, and activities.<sup>3</sup>

Ethics is “the discipline dealing with what is good and bad and with moral duty and obligation,” as well as “the principles of conduct governing an individual or a group.”<sup>4</sup> In commerce, an ethical mind-set supports values-based decision-making. The aim is to do not only what is good for business, but also what is good for the organization’s employees, clients, customers, and communities in which it operates.

Bringing together these two definitions, “AI ethics” refers to the organizational constructs that delineate right and wrong—think corporate values, policies, codes of ethics, and guiding principles applied to AI technologies. These constructs set goals and guidelines for AI throughout the product life cycle, from research and design, to build and train, to change and operate.

An ethical mind-set supports values-based decision-making—focusing on doing not only what is good for business, but what is good for an organization’s employees, clients, customers, and the communities in which they live and work.

# Considerations for carrying out AI ethics

Conceptually, AI ethics applies to both the goal of the AI solution, as well as each part of the AI solution. AI can be used to achieve an unethical business outcome, even though its parts— machine learning, deep learning, NLP, and/or computer vision— were all designed to operate in an ethical way. For example, an automated mortgage loan application system might include computer vision and other tools designed to read handwritten loan applications, analyze the information provided by the applicant, and make an underwriting decision based on parameters programmed into the solution. These technologies do not process such data through an ethical lens—they just process data. Yet if the mortgage company inadvertently programs the system with goals or parameters that discriminate unfairly based on race, gender, or certain geographic information, the system could be used to make discriminatory loan approvals or denials.

In contrast, an AI solution with an ethical purpose can include processes that lack integrity or accuracy toward this ethical end. For example, a company may deploy an AI system with machine learning capabilities to support the ethical goal of nondiscriminatory personnel recruiting processes. The company begins by using the AI capability to identify performance criteria based on the best performers in the organization’s past. Such a sample of past performers may include biases based on past hiring characteristics (including discriminatory criteria such as gender, race, or ethnicity) rather than simply performance. In other words, the machine learns based on the data that it processes, and if the data sample is not representative or accurate, then the lessons it learns from the data will not be accurate and may lead to unethical outcomes.

To understand where ethical issues could arise and how in the future of work those issues might be avoided, it helps to organize AI along four primary dimensions of concern (figure 1):

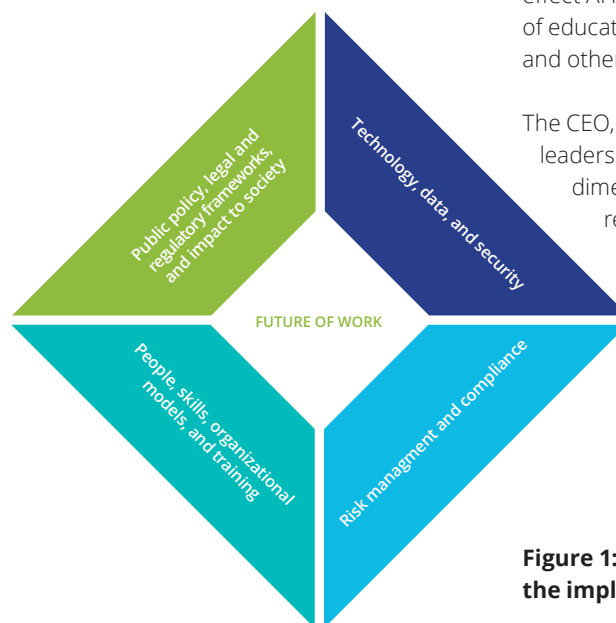
**Technology, data, and security.** Look at the organization’s approach to the AI life cycle from an ethical perspective, including the ways it builds and tests data and models into AI-enabled solutions. Leadership in this dimension comes from the organization’s information, technology, data, security, and privacy chiefs.

**Risk management and compliance.** Find out how the organization develops and enforces policies, procedures, and standards for AI solutions. See how they tie in with the organization’s mission, goals, and legal or regulatory requirements. The heads of risk, compliance, legal, and ethics play a role in this dimension.

**People, skills, organizational models, and training.** Understand and monitor how the use of AI impacts experiences of both employees and customers. Continuously assess how operating models, roles, and organizational models are evolving due to the use of AI. Educate all levels of the workforce and implement training initiatives to retool or up-skill capabilities. Establish protocols to incentivize ethical behavior and encourage ethical decisions along the AI life cycle. In this dimension, the human resources function shares responsibility with learning and development teams, ethics officers, and broader executive leadership.

**Public policy, legal, and regulatory frameworks, and impact to society.** Finally, develop a sense of AI’s place in the business environment. This includes the level of acceptance AI has in government and culture. It also includes the direction that laws and regulations are taking with regard to AI. Apply this information to the effect AI is likely to have over time in terms of education, employment, income, culture, and other aspects of society.

The CEO, CRO, CCO, and CFO have leadership roles across the first three dimensions, while the fourth dimension relies on leadership from politicians, regulatory agencies, and other policymaking bodies.



**Figure 1: Four dimensions are vital to the implementation of AI ethics**

### Examples of AI ethics risks

Ethical considerations exist along each stage of the AI life cycle.

Life cycle stage	Examples of risks	Tools for managing risk
<p><b>Research and design</b></p>	<p>The solution’s inherent risks (such as a computer vision application that captures and potentially misuses customers’ or employees’ images or other personally identifiable information [PII]) are overlooked during the design phase</p> <p>The solution’s intended purpose is not aligned to the organization’s mission and values</p>	<p>A framework for what defines ethical use of AI and data in the organization</p> <p>A cross-functional panel with representation from the business units as well as the technology, privacy, legal, ethics, and risk groups</p> <p>A data or AI ethics code of conduct by which professionals must abide</p>
<p><b>Build and train</b></p>	<p>The organization lacks appropriate ways to secure consent from individuals whose data is used to train the AI model</p> <p>A programmer builds bias (either consciously or unconsciously) into a model intended for an AI-enabled solution, such as in the personnel recruiting example previously described</p> <p>The data used to train an AI model has quality issues</p>	<p>A process for determining where and how to obtain the data that trains the models</p> <p>Guidelines on where and how user consent becomes a consideration in the training phase</p> <p>Policies for where and how to build models and whether to use open source technology</p> <p>An assessment of ways that an AI solution can teach itself behaviors out of sync with the organization’s mission or values</p>
<p><b>Change and operate</b></p>	<p>A chatbot (an AI application that can include cognitive language capabilities) learns new behaviors that are inappropriate or offensive to customers</p> <p>The organization is unable to quickly assess which new data sources an AI-enabled solution has recently accessed on its own</p> <p>The organization lacks the ability to test and monitor AI solutions</p>	<p>A process for the organization to engage in continuous monitoring</p> <p>An assessment of ways that an AI-enabled solution can gain access to new forms of data</p> <p>A process for the business to update the board on AI-related risks or issues</p> <p>Organizational thresholds or tolerance levels to help determine whether to decommission an AI-enabled solution</p>

# First steps along the AI ethics journey

AI ethics is a sweeping endeavor with many moving parts. At the same time, technology aside, the initial approach should follow a similar path as other ethics and compliance programs, including:

**Start at the top.** The board of directors and senior management set the tone for any ethics and compliance program. Inform top leadership of the use of AI tools across the organization and how AI presents opportunity and risk to the organization. Then, review and—wherever necessary—expand on the enterprise’s existing policies, procedures, and standards (or consider adopting new ones) to address AI ethics as a key priority of the organization.

**Clearly communicate your intentions.** Ethical constructs and mechanisms protect the organization while addressing risks associated with the use of data and technology. To begin with, consider developing an AI “code of conduct” for data scientists and other data professionals while setting up channels for escalating issues. Also consider establishing principles for the use of AI among management, employees, investors, and customers, as applicable to the organization.

**Assess the risks.** AI ethics is as much about understanding the risks as it is about establishing a process for avoiding them. Be clear on what kind of AI solution you are building and who you are building it for. Identify the processes in the AI life cycle that could negatively impact stakeholders, order them by priority, and allocate resources to mitigate the risks.

**Give examples.** Let product teams know what to look for in monitoring solutions for AI ethics. One way is to embed ethical use of data and technology considerations into IT architecture design (similar to the [Privacy by Design](#) framework). Another is to design control structures and embed them into AI-enabled solutions.

**Put up guardrails.** Organizations should consider proactively establishing guardrails to guide, monitor, and assess how AI is used by the organization, employees, vendors, and customers. Such guardrails may be either technical or organizational—akin to internal controls. An example of a technical guardrail is a control framework embedded within the design of an AI solution that prevents specific actions from being completed. Another technical guardrail example is the use of explainable and interpretable AI, where the decision-making behind AI-enabled solutions is both transparent and explainable. An example of an organizational guardrail is a cross-functional panel that assesses all AI-enabled solutions before they are built (similar to an [Institutional Review Board](#)) and considers the impact of emerging AI ethics-related issues on existing and as-yet implemented AI solutions. Such a panel should have accountability and a clear charter to drive change and influence decisions across the organization.

It is easy to get caught up in the complexity of AI. But starting with the basics can create near-term impact while offering maximum room to learn as you go. Over time, the organization can integrate the finer nuances of AI ethics as its implications—for the organization and stakeholders alike—become known.

## Learn more:

Visit the [Notre Dame Deloitte Center for Ethical Leadership](#)

The [Notre Dame Deloitte Center for Ethical Leadership \(NDDCEL\)](#)

is a collaboration between the University of Notre Dame and Deloitte.

Having identified a shared value of personal integrity in today’s business world, NDDCEL came together to advance the understanding and implementation of ethical leadership practices in the corporate sphere.

Today, with the increasing use of AI-enabled solutions, ethics and leadership are entering uncharted territory. The Center is currently focused on developing mutual understanding of ethical issues, challenges, opportunities, and solutions related to AI and analytics. As part of this initiative, the NDDCEL provides executives and business leaders with opportunities to learn more about this rapidly evolving topic, connect with like-minded peers, and help develop the next generation of business leaders by translating research insights into leading practices.

**i** To learn more, please visit the [NDDCEL’s website](#)

# Addressing the AI ethics imperative—everyone's responsibility

With its machine learning, deep learning, NLP, and computer vision capabilities, AI offers the exciting prospect of improving the human condition. But there is a potential dark side to AI that is hard to ignore.

The result is a new frontier in business ethics. Those involved with the advancement of AI—including corporate boards, management teams, researchers, and engineers—face a growing imperative to bring an ethical lens to what they design and build. This approach should be articulated through organizational ethical constructs that apply throughout the AI product

life cycle. To be effective, each construct should reflect an understanding of AI-related vulnerabilities along the previously described four dimensions of risk.

It is too soon to tell where this journey will lead businesses and their customers. But the way forward is in sight. It begins with a top-level commitment to ethical leadership and a focus on technical and ethical literacy. With ethical guardrails to limit missteps along the way, everyone in an organization can work together to produce solutions that represent AI's potential.





---

# Contact us

**Mary Galligan**

Managing Director  
Deloitte Risk & Financial Advisory  
Deloitte & Touche LLP  
mgalligan@deloitte.com

**Courtney Parry**

Senior Manager  
Deloitte Risk & Financial Advisory  
Deloitte & Touche LLP  
cparry@deloitte.com

**Vivek Katyal**

Principal  
Deloitte Risk & Financial Advisory  
Deloitte & Touche LLP  
vkatyal@deloitte.com

**Christopher Adkins**

Executive Director  
Notre Dame Deloitte Center for Ethical  
Leadership  
University of Notre Dame  
cadkins1@nd.edu

**Maureen Mohlenkamp**

Principal  
Deloitte Risk & Financial Advisory  
Deloitte & Touche LLP  
mmohlenkamp@deloitte.com

---

# Endnotes

1. "State of AI in the Enterprise, 2nd Edition," Loucks, Davenport, and Schatsky, Deloitte Insights, Oct. 22, 2018, <https://www2.deloitte.com/insights/us/en/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey.html>.
2. <https://www.merriam-webster.com/dictionary/artificial%20intelligence>.
3. Op cit, "State of AI in the Enterprise," 2nd Edition."
4. <https://www.merriam-webster.com/dictionary/ethic>.



This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

#### About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States, and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see [www.deloitte.com/about](http://www.deloitte.com/about) to learn more about our global network of member firms.