SYNERIO

Synerio Technologies

AI SAFETY, ETHICS AND RISK MANAGEMENT

Artificial Intelligence (AI) and Machine Learning (ML): Concepts, Applications, and Challenges

From the Center for AI Safety

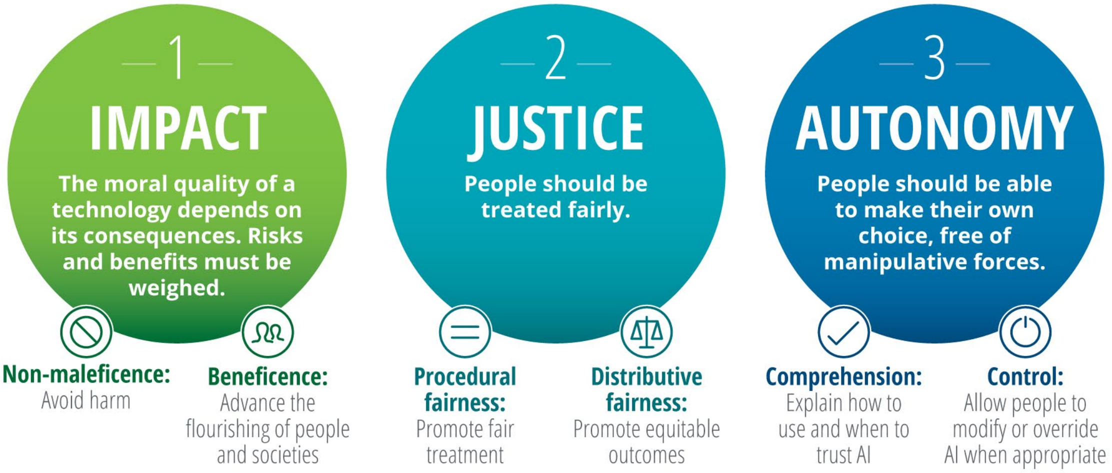
AI experts, journalists, policymakers, and the public are increasingly discussing a broad spectrum of important and urgent risks from AI. Even so, it can be difficult to voice concerns about some of advanced AI's most severe risks. The succinct statement below aims to overcome this obstacle and open discussions. It is also meant to create common knowledge of the growing number of experts and public figures who also take some of advanced AI's most severe risks seriously. From the use of Generative AI and Agentic AI which is becoming a popular concept as AI agents are being used autonomously by a master AI. Controls and monitoring in the realm of AI growth is imperative.

Mitigating the societal-scale risks such as pandemics and nuclear war

SYNERIO adopts a well thought out framework issued by Deloitte Insights regarding ethics, protections, and true purpose behind SYNERIO's use of AI in its products and services. In 2025, Synerio has included perspectives from Ernst & Young relating to some countries looking to regulate AI in areas related to bias of algorithms and copyrights. The U.S. remains unlikely to pass new legislation on AI any time soon.

FIGURE 1

Three core principles can help leaders think through AI's ethical implications


1
IMPACT
The moral quality of a technology depends on its consequences. Risks and benefits must be weighed.
Non-maleficence: Avoid harm
Beneficence: Advance the flourishing of people and societies
2
JUSTICE
People should be treated fairly.
Procedural fairness: Promote fair treatment
Distributive fairness: Promote equitable outcomes
3
AUTONOMY
People should be able to make their own choice, free of manipulative forces.
Comprehension: Explain how to use and when to trust AI
Control: Allow people to modify or override AI when appropriate

Source: Deloitte analysis.

Deloitte Insights | deloitte.com/insights

Impact: Promoting acceptable outcomes

Two widely recognized ethical principles are *non-maleficence* (“do no harm”) and *beneficence* (“do only good”). These principles are grounded in “consequentialist” ethical theory which holds that the moral quality of an action depends on its consequences.

“First, do no harm”

Non-maleficence prescribes that AI should avoid causing both foreseeable and unintentional harm. Examples of the former could include weaponized AI, the use of AI in cyberwarfare, malicious hacking, the creation or dissemination of phony news or images to disrupt elections, and scams involving phishing and fraud. But of course, the great majority of organizations building or deploying AI have no intention of causing needless harm. For them, avoiding unintended consequences is the paramount concern.

Avoiding harmful AI requires that one understand AI technologies’ scientific limitations to manage the attendant risks. For example, many AI algorithms are created by applying machine learning techniques, most prominently deep learning, to large bodies of “labeled” data. The resulting algorithms can then be deployed to make predictions about future cases for which the true values are unknown.

Knowing that machine learning algorithms perform reliably only to the extent that the data used to train them suitably represents the scenarios in which they are deployed, an organization can take steps to identify and mitigate the risks arising from this limitation.

AI for good

The principle of ***beneficence***, reflected in many AI ethics declarations, holds that AI should be designed to help promote the well-being of people and the planet.

While non-maleficence is a common principle in the AI ethics declarations, the principle of beneficence appeared in less than half of those declarations. It is possible that this disparity reflects a prevalence of alarmist discussions of AI that focus more on harm, but dwell less on AI’s potential to help debias human decisions, extend human capabilities, and improve well-being.

Managing tradeoffs

Ethical deliberations often involve managing tradeoffs between different principles that cannot be simultaneously satisfied. Tradeoffs between beneficence and non-maleficence are common. Sometimes, the process of articulating an ethical tradeoff can spur innovations that render the tradeoff less fraught.

The broader point is that ethical AI requires organizations to consider not only *predictions*, but *interventions* as well. The newer science of choice architecture expands the toolkit with “soft” interventions that can allow organizations to act ethically on ambiguous algorithmic indications. In cases where nudge interventions aren’t strong enough, ethical deliberation should help guide policy decisions about how machine-generated predictions are acted upon.

A still broader point is that technological innovation, often involving multidisciplinary thinking, can also make it possible to mitigate difficult ethical tradeoffs. The increasingly popular tagline “human-centered AI” can perhaps be interpreted as a call to take human and societal needs into account when developing uses for AI technologies.

Justice: Treating people fairly

Justice is another core ethical principle that appears frequently in AI ethics declarations. It encompasses such related concepts as inclusion, equality, diversity, reversibility, redress, challenge, access and distribution, shared benefits, and shared prosperity.

Much of the conversation about justice as it relates to AI revolves around “algorithmic fairness”—the idea that AI algorithms should be fair, unbiased, and treat people equally. But what does it mean for an algorithm to be “fair”?

It is useful to distinguish between the concepts of procedural and distributive fairness. A policy (or an algorithm) is said to be ***procedurally*** fair if it is fair independently of the outcomes it produces. Procedural fairness is related to the legal concept of due process. A policy (or an algorithm) is said to be ***distributively*** fair if it produces fair outcomes. Most ethicists take a distributive view of justice, whereas a procedure’s fairness rests largely on the outcomes it produces.

Ethical AI by design

Ethics is often viewed as a ***constraint*** on organizations' abilities to maximize return on innovation. Ethical principles can serve as design criteria for developing innovative uses of AI that can improve well-being, reduce inequities, and help individuals better achieve their goals. In this sense, the principles of impact and justice can help shape AI technologies in ways that achieve what marketing, management, and design professionals, respectively, call customer-centricity, employee-centricity, and human-centricity. Developing trustworthy AI technologies that safely and fairly help advance these goals is a distinctly 21st-century way for organizations to do well by doing good.

Trust and Transparency

The purpose of AI is to augment human intelligence. 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. Our customer's 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.

Closing Statement

In embracing the limitless potential of artificial intelligence, we recognize that trust and transparency are the cornerstones of responsible innovation. As stewards of this powerful technology, we pledge to uphold the highest ethical standards, ensuring that our AI systems are guided by fairness, accountability, and a commitment to the well-being of all. Through unwavering transparency, we invite you to join us on this journey, fostering a future where AI serves humanity with empathy, integrity, and a profound sense of responsibility. Together, we can build a future where innovation and ethics go hand in hand, inspiring trust and shaping a world that reflects our shared values**. While it is wise to watch what other countries are doing from a risk perspective and regulations to protect their citizens, there may be some regulations and protocols that could make sense to implement to offer our customer community the safest AI possible.**