

AI for social good: Improving lives and protecting the planet

*This report is a collaborative effort by
Medha Bankhwal, Ankit Bisht, Michael Chui,
Roger Roberts, and Ashley van Heteren,
representing views from McKinsey Digital.*



AI is already being used to further all 17 SDGs—from the goal of eliminating poverty to establishing sustainable cities and communities and providing quality education for all—and generative AI has opened new possibilities. As we look to the future, we see exciting potential for acceleration, with new tools and platforms putting ever-greater power in the hands of social entrepreneurs, public sector innovators, and private sector players to create effective solutions. But that power also brings with it the need to assure it is harnessed in trusted and responsible ways and that risks are monitored and managed actively to avoid unintended harms.

Six years ago, it was becoming clear that AI could play a major role globally in promoting not just productivity and economic growth but also social good. In a 2018 report, we outlined how AI capabilities, from natural language processing to sound recognition and tracking, could be used in about 170 use cases to benefit society¹—to promote equality and inclusion, improve crisis response, protect the environment, and deliver impact in many more ways.²

Today's AI R&D is not just confirming our initial assessments but showing promise for further gains in the future. A series of improvements in AI techniques and progress on key enablers have substantially expanded the universe of problems that AI may be able to address. Much of this progress is centered on generative AI, which is enabling natural language interfaces; rapid language translation; synthesis of vast document repositories; creation of stories in text, images, and video; and much more.³

In this report, we take another look at how AI can become a key part of solutions to benefit people and the planet—and how it already has. One way to assess this is by mapping innovations and impact to the UN Sustainable Development Goals, or SDGs (see sidebar “Methodology,” found at the end of the report). The SDGs comprise 17 goals and 169 targets that aim to improve lives around the world and protect the planet. But the UN's 2023 update on progress toward the SDGs indicates the world is on track to meet only 15 percent of SDG targets.⁴ In real terms, this means that 2.2 billion people lack access to safe water and hygiene, and 3.5 billion lack access to safely managed sanitation⁵; roughly 3.3 billion people live in environments that are highly vulnerable to climate change⁶; and about 750 million people are facing hunger.⁷

Below, we illustrate the potential of AI to catalyze progress on these pressing social issues, and we highlight the challenges in the domains of data quality and governance, as well as access to AI talent (particularly for not-for-profits), that are hindering AI from scaling. We then outline some actions that stakeholders—including governments, foundations, universities, and businesses—could take to overcome these challenges. While the opportunities have associated risks, such as embedded biases and data privacy and security threats, thoughtful action could accelerate the deployment of AI-based solutions to advance progress on the SDGs and improve lives across the globe.

¹ “Applying artificial intelligence for social good,” McKinsey Global Institute, November 28, 2018. Additional research in 2023 yielded discovery of 13 more use cases piloted in 2018 that were not originally accounted for in our 2018 report, bringing the 2018 total up to about 170.

² “Tech for Good: Using technology to smooth disruption and improve well-being,” McKinsey Global Institute, May 15, 2019; Amine Aït-Si-Selmi, Eric Hazan, Hamza Khan, and Tunde Olanrewaju, “Tech for Good: Helping the United Kingdom improve lives and livelihoods,” McKinsey, July 31, 2020.

³ For more on generative AI, see “What is generative AI?,” McKinsey, April 2, 2024.

⁴ *The Sustainable Development Goals report 2023: Special edition*, United Nations, July 10, 2023.

⁵ “The 17 goals,” United Nations Department of Economic and Social Affairs, accessed April 24, 2024.

⁶ “Protecting people from a changing climate: The case for resilience,” McKinsey, November 8, 2021.

⁷ “122 million more people pushed into hunger since 2019 due to multiple crises, reveals UN report,” World Health Organization, July 12, 2023.

AI is not a magic bullet, and many risks need to be managed to harness its potential (see sidebar “Managing the risks of adopting AI”). But the universe of problems that AI can address is broad.

Current applications of AI are applicable to all the SDGs, including modeling proteins, screening drugs, designing vaccines, targeting aid and public services, solving supply chain problems such as route optimization for last mile delivery of food in remote geographies, forecasting the long-term impacts of climate change or giving early warning for natural disasters, and bringing expertise to frontline aid workers.

Additionally, adoption of generative AI could significantly increase and democratize access to new capabilities. AI tools now allow remote users to complete tasks that once required specific expertise, such as language translation, fact checking, identification of human or plant diseases, and identification of harmful online content. In a recent survey of more than 4,000 not-for-profits conducted by Google for Nonprofits, 75 percent of respondents said that generative AI had the potential to transform their marketing efforts by enhancing their translation and fact-checking capabilities.⁸

The experts we interviewed noted that AI could address or help solve social or environmental challenges in two circumstances: 1) when the AI solution could solve problems that bottleneck other efforts in the field—for example, a solution for water leakage in residential pipes requires predictions about the likelihood of leaks based on analysis of data such as pipe age and location; and 2) when data required for the model to work is (or will soon be) available and accessible.

To map the breadth of AI’s applicability, we have developed a database of AI use cases, each of which highlights a type of meaningful problem whose solution could be enabled by one or more AI capabilities. At the time of our 2018 report, this database contained about 170 high-potential use cases. It now contains about 600—more than a threefold increase. This number is growing as more innovative uses come to light, as social impact leaders continue to experiment boldly, and as AI tools become more accessible and easier to use.

The number of real-life AI deployments has also increased significantly over the past six years. In 2018, only a small fraction of the about 170 use cases had been deployed. Today, about 490 of the 600 use cases, or more than 80 percent, have been implemented in at least one instance (Exhibit 1).⁹

Adoption of generative AI could significantly increase and democratize access to new capabilities. AI tools now allow remote users to complete tasks that once required specific expertise.

⁸ *The Keyword*, “3 insights from nonprofits about generative AI,” blog entry by Annie Lewin, March 28, 2024.

⁹ Our library contains approximately 600 use cases, and our analysis of deployments is based on publicly available data. Neither is comprehensive or exhaustive, and both will continue to evolve.

Additionally, several SDGs that are behind on progress have relatively untapped AI potential. Consider the following examples:

- No Poverty (SDG 1): machine learning could be used to direct cash aid to those most in need or provide alternative credit scores to financially excluded individuals.
- Zero Hunger (SDG 2): AI could be used to help develop new crops, better select crop regions to minimize crop risks, and provide early warning for nutrition crises.
- Peace, Justice, and Strong Institutions (SDG 16): machine learning could be used to detect and curb the spread of misinformation, provide access to information that enables advocacy for policy change, and improve measurement of specific policy interventions.

Below, we explore potential and existing deployments in three of the SDGs with the most widely recognized potential: Good Health and Well-Being (SDG 3), Quality Education (SDG 4), and Climate Action (SDG 13). We also explore two SDGs where AI does not have widely recognized potential but has had an impact in select areas: No Poverty (SDG 1) and Zero Hunger (SDG 2).

We explore potential and existing deployments in three of the SDGs with the most widely recognized potential. We also explore two SDGs where AI does not have widely recognized potential but has had an impact in select areas.

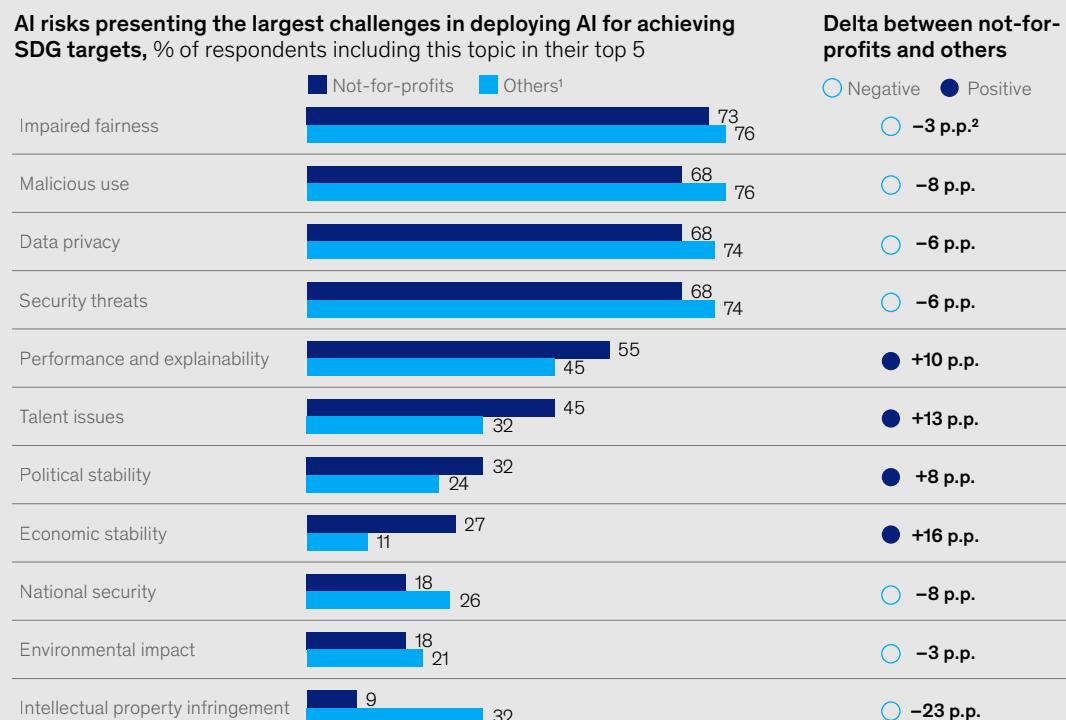
Managing the risks of adopting AI

Risks are inherent to the use of AI. With generative AI (gen AI), risks include inaccurate outputs, biases embedded in the underlying training data, the potential for large-scale misinformation, and malicious influence on politics and personal well-being.¹ As we have noted in multiple recent articles,² AI tools and techniques can be misused, even if they were originally designed for social good.

Respondents to our survey of about 60 experts identified the top risks as impaired fairness, malicious use, and privacy and security concerns, followed by explainability (exhibit).³ Respondents from not-for-profits expressed relatively more concern about misinformation, talent issues such as job displacement, and effects of AI on economic stability compared with their counterparts at for-profits, who were more often concerned with intellectual property infringement.

Exhibit

Experts say impaired fairness and malicious use are the top risks in using AI to address the Sustainable Development Goals.



Note: "Impaired fairness" was framed as "bias and fairness" in the survey; "performance and explainability" was framed as "explainability"; "data privacy" and "security threats" were combined in the survey.

¹"Others" includes for-profits, think tanks, academic institutions, and consultancies.

²Percentage points.

Source: Survey of ~60 experts representing 48 organizations (incl not-for-profits, foundations, technology companies, start-ups, academic institutions, and government) and 17 countries

McKinsey & Company

¹ "Implementing generative AI with speed and safety," *McKinsey Quarterly*, March 13, 2024.

² Ibid.; "The state of AI in 2023: Generative AI's breakout year," McKinsey, August 1, 2023; *New at McKinsey Blog*, "An inside look at how businesses are—or are not—managing AI risk," blog entry by Liz Grennan and Bryce Hall, August 31, 2023; "What is generative AI?," McKinsey, April 2, 2024.

³ Our AI risks framework for social impact builds on McKinsey's gen AI risks framework (see "Implementing generative AI with speed and safety," *McKinsey Quarterly*, March 13, 2024). It includes additional categories such as political stability and environmental impact and excludes risks such as strategic risks that can be more relevant to for-profit enterprises.

Impaired fairness. Algorithmic systems can inherit biases from their creators or from the data sets on which they are trained. When these algorithms are deployed in decision-making capabilities, these biases can reinforce preexisting prejudices and social inequalities, with potentially negative impacts on marginalized communities. One organization, Data Science for Social Good, builds bias detection tools that allow developers to audit data science systems for bias and equity.⁴

Malicious use. Malicious use includes creating and disseminating false information or fake content, scams, phishing attempts, hate speech, and activities that harm individuals and national security. A 2022 UN report found that misinformation had been used to incite hatred against marginalized groups and to prevent civilians from finding humanitarian corridors during conflicts, such as the one in Ukraine.⁵ According to a recent report by the World Economic Forum, “growing misinformation and disinformation could further increase vaccine hesitancy, which has already led to the re-emergence of locally eradicated diseases.”⁶ The Global Disinformation Index uses models based on large language models (LLMs) to detect disinformation with the goal of tracking news sites supported by hostile states.⁷ Similarly, Full Fact is an independent fact checking organization that deploys a range of AI and machine learning methodologies to detect and curb the proliferation of misinformation across the evolving landscape of information-spreading platforms.

Data privacy and security threats. Many of the UN Sustainable Development Goals use cases require access to health or financial data of vulnerable populations. While organizations are well aware of the harm that could result from breaches in their data systems, many social enterprises have resource constraints that may limit their ability to use the latest cybersecurity capabilities. Several organizations have developed data privacy guidelines, tool lists, and custom security frameworks for not-for-profits with limited resources.⁸

Performance and explainability. Many AI solutions employ complex algorithms that can make it difficult to identify the data or logic used to arrive at a decision. This is particularly relevant for gen AI solutions, which may provide inaccurate or toxic answers. Explainable AI models have several advantages for not-for-profits: they may make it easier to verify the correctness and fairness of results, to assign credit to data providers, and to assign accountability for model outcomes. The Allen Institute for AI recently released a platform for comparing large text data sets to measure the prevalence of toxic, low-quality, duplicate, or personally identifiable information used to train various LLMs.⁹

To mitigate the risks of AI, organizations must first understand and prioritize the risks they are most likely to face, both from inbound AI threats such as disinformation and from developing and deploying their AI solutions. While risks such as data privacy may be addressed through traditional software tools, emerging risks, such as bias in systems driven by LLMs, may require the development of new monitoring systems and guardrails.

⁴ “The bias and fairness audit toolkit for machine learning: Aequitas,” Center for Data Science and Public Policy, accessed April 24, 2024.

⁵ A/77/288: *Disinformation and freedom of opinion and expression during armed conflicts - Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, Office of the High Commissioner for Human Rights, United Nations, August 12, 2022.

⁶ *The global risks report 2023: 18th edition*, World Economic Forum, 2023.

⁷ “What we do,” Global Disinformation Index, accessed April 24, 2024.

⁸ “Online privacy for nonprofits: A guide to better practices,” Electronic Frontier Foundation, accessed April 24, 2024; “Learn,” Digital Defense Fund, accessed April 24, 2024; website of SOAP, accessed April 24, 2024; “Frontline policies,” Open Briefing Ltd, accessed April 24, 2024.

⁹ Akshita Bhagia et al., “What’s in my big data?”, arXiv:2310.20707, March 2024.