Chapter 8: Societal Considerations
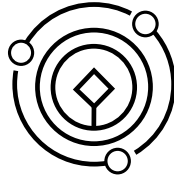
Introduction

This chapter begins by identifying the topics in ethical challenges mentioned in 59 Ethical AI Principle documents based on a dataset compiled by PricewaterhouseCoopers (PwC). The chapter also documents the key topics discussed in global news media on AI and Ethics based on LexisNexis data and Quid. AI use cases supporting each of the 17 United Nations (UN) Sustainable Development Goals (SDGs) are identified based on curated data from the McKinsey Global Institute (MGI).

Ethical Challenges

AI systems raise a broad variety of ethical challenges that are now the concern of government, public interest organizations, NGO's, academia, and industry. Efforts to identify these challenges and to develop guiding principles for ethically and socially responsible AI systems are emerging from each of these sectors,. This snapshot of some such t efforts was derived from an analysis of more than 100 documents.

PricewaterhouseCoopers (PwC) compiled a dataset of ethical challenges (based on topic modeling) by looking at ethical AI guidelines across for 110 documents, of which only 59 were deemed to discuss a set of AI principles. Many were simply reviews or recommendations, and were not included in the analysis. The list of organizational documents and the list of principles is available in the Appendix.

A view of ethical AI frameworks over time is plotted identifying Associations and Consortiums, Industry and Consultancy groups, Governments, Tech Companies, and Think Tanks/Policy Institutes and Academia (Figure 8.1a). It is interesting to note that initial impetus for Ethical Principles sprang from Associations and Consortiums, with other organizations subsequently releasing their respective AI Principles in 2018 and 2019.

Top 3 Ethical Challenges, Associations and Consortiums, Governments, and Tech Companies

Associations and Consortiums (19 documents)
1.) Interpretability & Explainability is cited in 95% of frameworks.
2.) Fairness is cited in 89% of frameworks.
3.) Transparency is cited in 84% of frameworks.
Governments (13 documents)
1.) Interpretability & Explainability, Fairness, and Transparency are each cited in 92% of frameworks..
Tech Companies (11 documents)
1.) Fairness is cited in 100% of frameworks.
2.) Transparency is cited in 81% of frameworks.
3.) Accountability is cited in 72% of frameworks.
Think Tanks/Policy Institutes and Academia (8 documents)
1.) Fairness is cited in 100% of frameworks.
2.) Human Control is cited in 88% of frameworks.
3.) Interpretable & Explainable Model is cited in 88% of frameworks.
Industry and Consultancy (8 documents)
1.) Transparency is cited in 88% of frameworks.
2.) Fairness, Data Privacy, and Reliability, Robustness, and Security are each cited in 75% of frameworks.

Number of Ethical AI Frameworks Produced 2016-2019, by Type of Organization

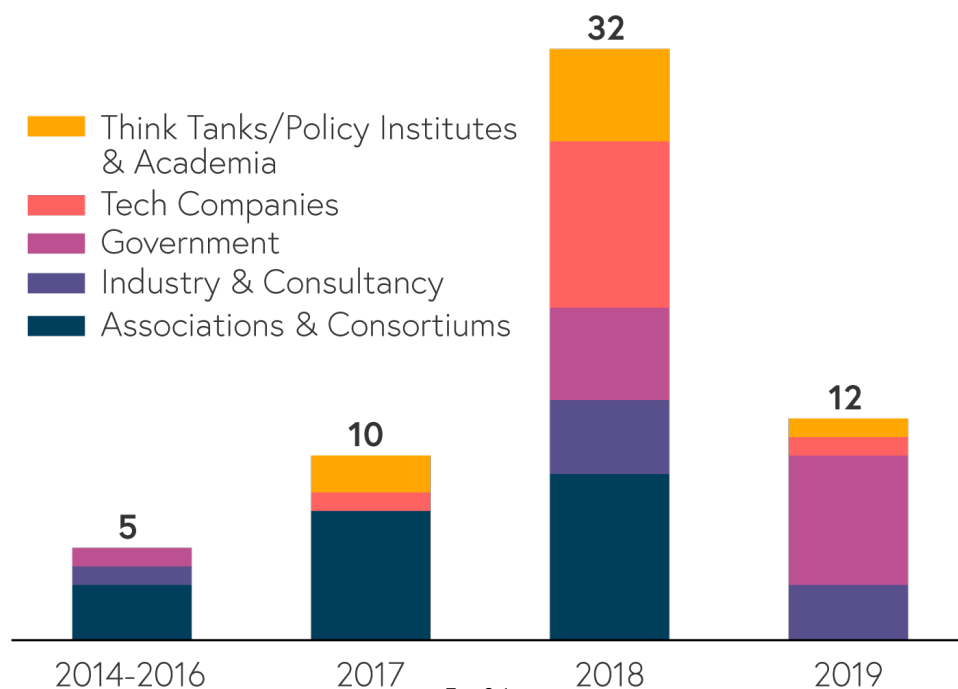Source: PwC based on 59 Ethical AI Principle documents.

Fig 8.1a.

Ethical Challenges

Twelve ethical challenges were mentioned across many ethical AI framework documents. This list is non-exhaustive, and many important ethical issues -- including justice, economic development, poverty reduction, and inequality, are missing. Even so, these 12 ethical challenges indicate where attention has been focused:

- Accountability
- Safety
- Human Control
- Reliability, Robustness, and Security
- Fairness
- Diversity and Inclusion
- Sustainability
- Transparency
- Interpretability and Explainability
- Multi Stakeholder engagement
- Lawfulness and Compliance
- Data Privacy

To communicate the thrust of the ethical AI issues to the general public, the bar graph shows the incidence of identified ethical challenges across 59 AI Principles documents (Figure 8.1b). It shows that Fairness, Interpretability and Explainability, Transparency are most mentioned across all documents studied.

Ethical Challenges covered across AI Principle Documents

Source: PwC based on 59 Ethical AI Principle documents.

Fig 8.1b.

"Research around Ethical AI, especially on fairness, accountability, and transparency (FAT) of machine learning models has grown significantly in the past couple of years. While there is a broad consensus emerging on the core set of principles associated with ethics and AI, the contextualization of these principles for specific industry sectors and functional areas is still in its infancy. We need to translate these principles into specific policies, procedures, and checklists to make it really useful and actionable for enterprise adoption."
Anand Rao, Global AI Lead, PwC

Ethics and AI: Global News Media

Global news coverage of Artificial Intelligence has increasingly shifted toward discussions about its ethical use. To better understand how these narratives are taking shape, we leveraged Quid to search the archived news database of LexisNexis for news articles from 60,000 global English news sources and over 500,000 blogs on AI ethics from August 12, 2018 to August 12, 2019 (see Appendix for more detail on search terms).

Based on keywords defined by Harvard (seen here), Quid included search terms such as human rights, human values, responsibility, human control, fairness, discrimination or non-discrimination, transparency, explainability, safety and security, accountability, and privacy related to AI technology. Then, we selected the 10,000 most relevant articles using the platform's NLP algorithm and visualized unique articles.

Each node (or dot) on a Quid network map represents a single news article. Links connecting these articles denote articles that share similar language. When a large number of similar articles are identified and linked, clusters form to reveal unique topics. The Quid algorithm classified the resulting media narratives into seven large themes based on language similarity: **Framework and Guidelines** (32%), **Data Privacy Issues** (14%), **Facial Recognition** (13%), **Algorithm Bias** (11%), **Big Tech Advisory on Tech Ethics** (11%), **Ethics in Robotics and Driverless Cars** (9%), and **AI Transparency** (6.7%).

Quid network with 3,661 news articles on AI Ethics from August 12, 2018 to August 12, 2019. Colored by theme. Labeled by theme.
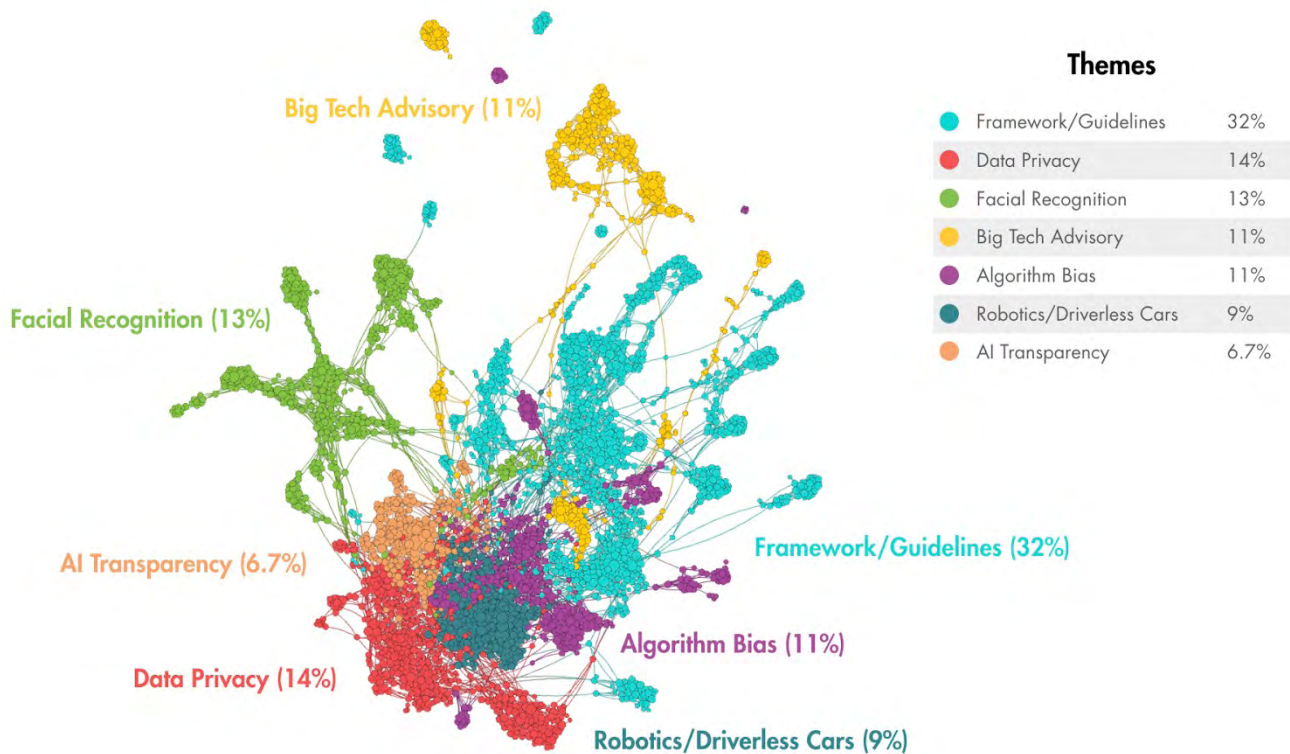

Fig. 8.2a.
Appendix: How to Red a Quid Network

Ethics and AI: Global News Media

These results indicate that the global media conversation on AI Ethics in 2019 is largely about AI ethics frameworks or guidelines led by governments, intergovernmental organizations, and research institutes (Figure 8.2a). Within the last year, nearly a third (32%) of all news articles covered AI guidelines proposed by governments or other large policy institutes, including those by the European Union and the Organisation for Economic Co-operation and Development (OECD). A smaller, but not an insignificant chunk of the conversation (11%) also included commentary from advisory groups attached to tech giants such as Google, Facebook, and Microsoft.

When filtering for ethics discussions around specific AI technologies, facial recognition dominated the attention of the news media, with 13% of all articles (Figure 8.2a). his cluster's position on the periphery of the larger AI ethics narrative indicates a high degree of uniqueness from the rest of the conversation. Public concerns over the technology's threat to data privacy have grown over time, driven by news of mistaken identities during crime surveillance, biometric scans that can be applied to videos or photos without consent, and the idea of data ownership as it relates to social media platforms that utilize the technology.

Countries differ significantly with respect to which AI ethical issues (as defined by Harvard here) they give most news coverage. While media sources based in the US or UK had more balanced coverage between categories, others reflected specific focus areas (Figure 8.2b). In Switzerland, for example, 45% of all articles covered guidelines and frameworks on AI development, while 44% of Chinese news focused on safety and security, and 48% of articles in Singaporean sources explored transparency and explainability.

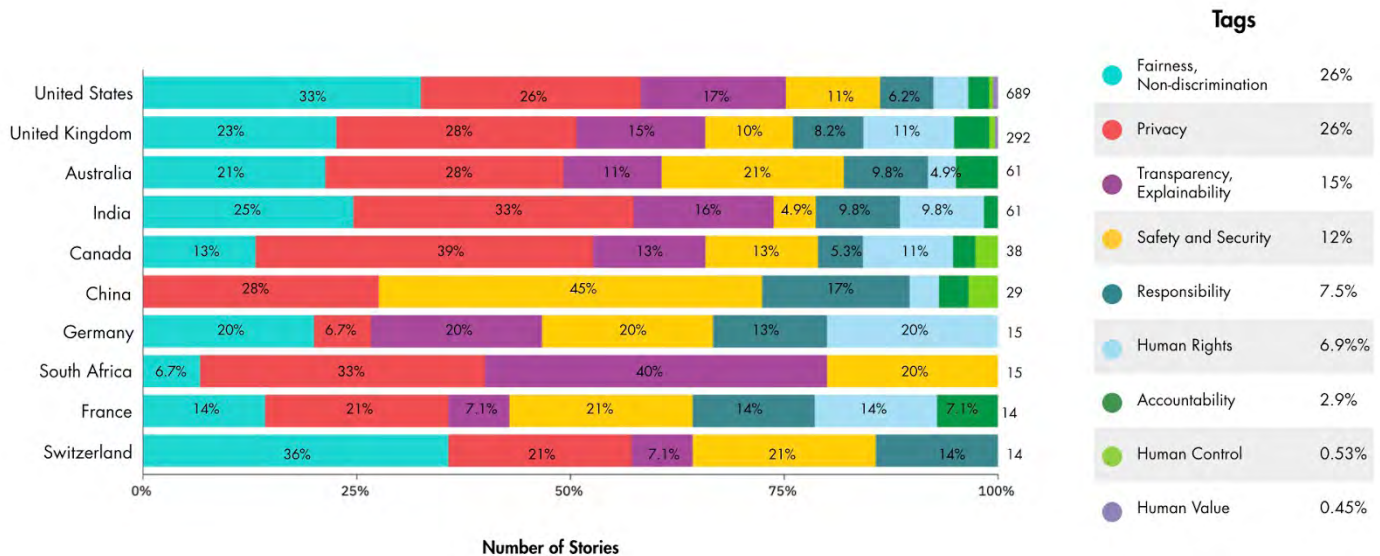### Most mentioned ethics categories by Source Country

Fig. 8.3b.

Applications of AI for Sustainable Development

Artificial intelligence, while not a silver bullet, has the potential to help contribute to multi-pronged efforts to address some of society's most pressing challenges.

The mapping of AI use cases to the UN Sustainable Development Goals (SDGs) that follows are derived from a library of approximately 160 AI for social good use cases collected by the McKinsey Global Institute and Noble Intelligence, McKinsey's initiative to use AI for humanitarian purposes. The library of use cases is not comprehensive, but reflects a selection of use cases, typically in domains with initial evidence of possible applications. AI deployments in some form were identified for about one-third of use cases in the library; in about three-quarters of use cases, deployments of solutions employing some level of advanced analytics were observed, most (if not all) of which could further benefit from using AI.

To build the use case library, MGI took a two-pronged approach: from a societal point of view, MGI sought to identify key problems known to the social sector community and determine where AI could aid efforts to resolve them; from a technological point of view, MGI took a curated list of 18 AI capabilities and sought to identify which types of social problems they could best contribute to solving. Each use case highlights a meaningful problem that can be solved by an AI capability or some combination of AI capabilities. The library is not comprehensive, but it nonetheless showcases a wide range of problems where AI can be applied for social good. MGI's full discussion paper can be found at Notes from the AI frontier: Applying AI for social good.

Applications of AI for Sustainable Development

Artificial intelligence has applicability across all 17 of the United Nations Sustainable Development Goals

The UN SDGs are a collection of 17 global goals set by the United Nations for the year 2030, for poverty alleviation, improving health and education, reducing inequality, preserving the environment, and boosting economic growth, amongst other priorities. AI use cases have the potential to support some aspect of each of the UN SDGs. The chart below indicates the number of AI use cases in MGI's library that could support each of the UN SDGs (Figure 8.3a).

SDG 3, "Ensure healthy lives and promote well-being for all at all ages", could be supported by the highest number of use cases in MGI's current library. A number of use cases that leverage AI support medical diagnoses: for example, researchers at the University of Heidelberg and Stanford University have created an AI system to visually diagnose skin cancer that outperformed professional dermatologists. There are also potential cases where AI can be

used to monitor, track and predict outbreaks of communicable diseases. For instance, Data Science for Social Good and McKinsey's Noble Intelligence initiative developed an algorithm to identify children most at risk of not receiving the measles vaccination, allowing physicians to spend more time educating and following up with these families.

There are also a number of AI use cases that could support SDG 16, "Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels." The use cases cover domains ranging from helping individuals verify and validate information, providing improved security through detection and prediction of violence, addressing bias to ensure fair and equal access to justice, to optimizing the management of public and social sector institutions. For example, AI could be used to automate question response or provision of services through digital channels, helping to improve government interactions with citizens.

AI use cases that support the UN Sustainable Development Goals
Source: 'Notes from the AI Frontier: Applying AI for social good', McKinsey Global Institute

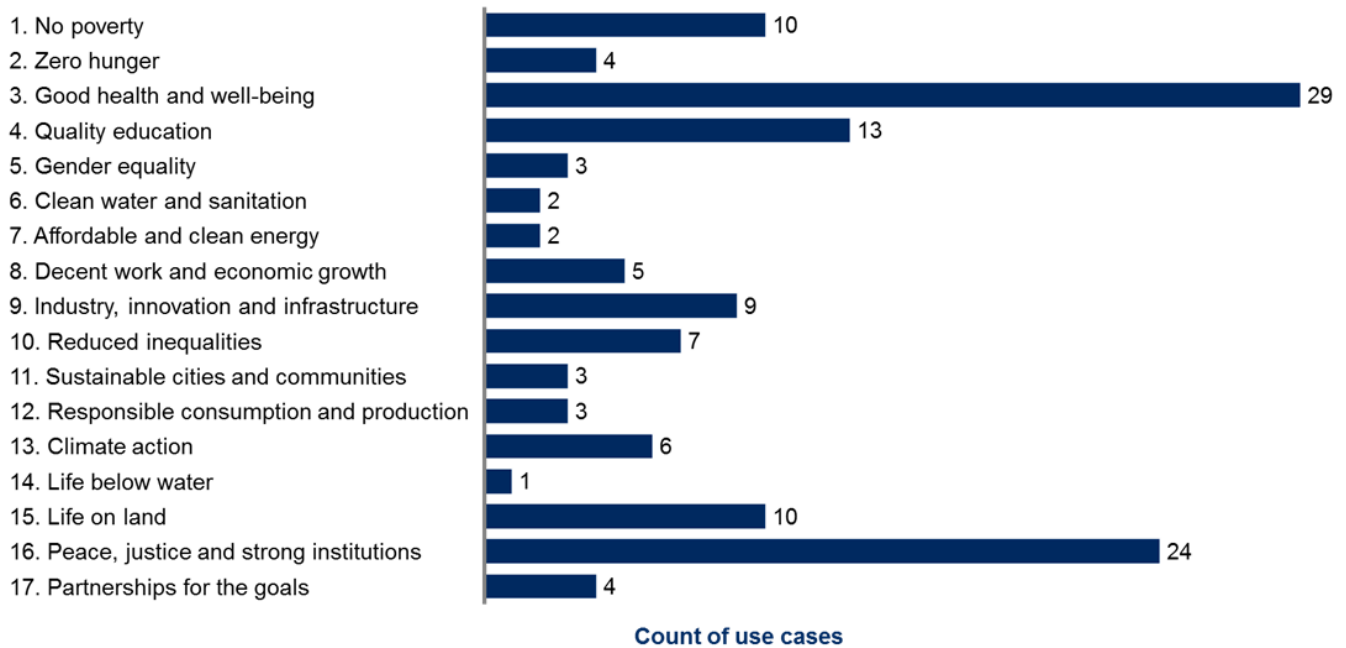| SDG	Count of use cases
1. No poverty	10
2. Zero hunger	4
3. Good health and well-being	29
4. Quality education	13
5. Gender equality	3
6. Clean water and sanitation	2
7. Affordable and clean energy	2
8. Decent work and economic growth	5
9. Industry, innovation and infrastructure	9
10. Reduced inequalities	7
11. Sustainable cities and communities	3
12. Responsible consumption and production	3
13. Climate action	6
14. Life below water	1
15. Life on land	10
16. Peace, justice and strong institutions	24
17. Partnerships for the goals	4

Count of use cases

Fig. 8.3a.
NOTE: This chart reflects the number and distribution of use cases and should not be read as a comprehensive evaluation of AI potential for each SDG; if an SDG has a low number of cases, that is a reflection of our library rather than of AI applicability to that SDG.

Applications of AI for Sustainable Development

AI is applicable to driving a subset of targets across the UN SDGs

Each UN SDG is broken down into a list of targets, which are measured with indicators. There are 169 targets across the 17 UN SDGs. While AI use cases can be topically aligned to the SDGs, as displayed in the previous chart, further focus should be directed to the use cases that can directly drive impact towards achieving specific UN SDG targets and indicators.

By mapping AI use cases to the specific target(s) that they could contribute to achieving, MGI identified the subset of targets for which AI has some applicability to address. This analysis builds upon the ~160 use cases in MGI's library and others to identify which targets could be addressed by a solution in which AI is applied, recognizing that AI alone cannot solve any of the targets. The following chart displays the number of targets which AI could contribute to addressing, out of the total number of targets within each SDG (Figure 8.3b).

Some AI for sustainable development use cases are being piloted, although bottlenecks exist

A number of organizations globally are piloting applications of AI for sustainable development, although there are currently few examples of deployments of AI for sustainable development at scale. For example, AI has been piloted for several applications in disaster relief by a number of organizations, including Google, Facebook, Microsoft, Planet Labs, Airbus, SAP, and others. Still, there is more to be done to sustainably adopt these AI applications for widespread use in disaster relief across multiple partners and regions.

Some AI-specific bottlenecks will need to be overcome for AI to reach its potential for social impact. These range from challenges with data (including availability, accessibility, quality, volume, labelling, and integration), accessing to computing capacity, availability and accessibility of AI talent, and the receptiveness and capabilities of organizations deploying solutions. Some efforts are underway to address this, especially to address accessibility of data for social good, including the Global Data Commons and UN Global Pulse.

AI applicability to address UN SDG targets
Source: UN Global Indicator Framework, McKinsey Global Institute analysis

■ Targets that could be supported by AI
□ Total number of targets

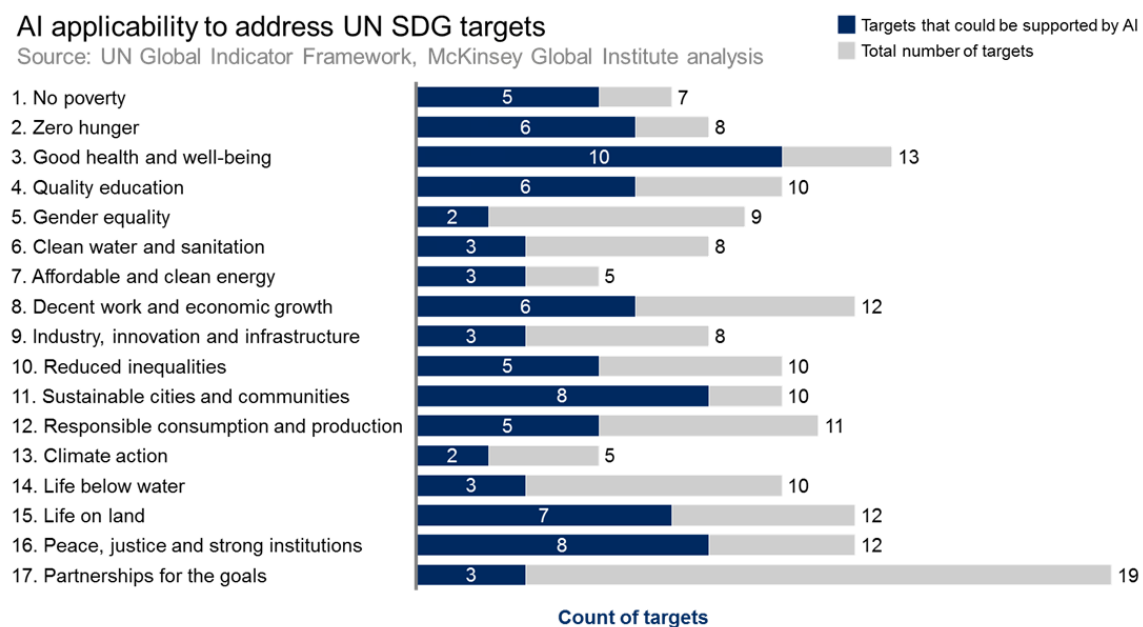| SDG	Targets that could be supported by AI	Total number of targets
1. No poverty	5	7
2. Zero hunger	6	8
3. Good health and well-being	10	13
4. Quality education	6	10
5. Gender equality	2	9
6. Clean water and sanitation	3	8
7. Affordable and clean energy	3	5
8. Decent work and economic growth	6	12
9. Industry, innovation and infrastructure	3	8
10. Reduced inequalities	5	10
11. Sustainable cities and communities	8	10
12. Responsible consumption and production	5	11
13. Climate action	2	5
14. Life below water	3	10
15. Life on land	7	12
16. Peace, justice and strong institutions	8	12
17. Partnerships for the goals	3	19

Count of targets

Fig. 8.3b.

Measurement Questions

- How can standardized granular data on AI use cases that impact fairness, human rights, and human dignity be generated?

- How can AI development be integrated into frameworks with social goals, to better plan AI technical development alongside social impacts?

- What measurements can be developed to assess how AI might generate societal threats as well as opportunities?