

# Ethics of Artificial Intelligence (AI)

Posted at: 07/02/2024

## Context:

1. Artificial Intelligence (AI) is being deployed in ways that touch people's lives, including in areas of healthcare, financial transactions, and delivery of justice.
2. Advances in AI can have profound impacts across varied societal domains, and in recent years, this realisation has sparked ample debate about the values that should guide its development and use.
3. Hence, there is a need to resolve differences in interpretation or attribution and document a better articulated AI landscape.

## Five ethical principles of AI:

1. Transparency: transparent processes in the development and design of AI algorithms, to increase interpretability, explain ability, or other acts of disclosure.
2. Justice and fairness: AI may increase inequality and reinforce societal biases if they are not addressed adequately.
3. Responsibility and accountability: This include clarifying legal liability, focusing on underlying processes that may cause potential harm, or whistleblowing in case of potential harm.
4. Privacy: Both as a value to uphold and as a right to be protected in ethical AI, in relation to data protection and data security
5. Non-maleficence: It indicates the precedence of moral obligation to preventing harm over the promotion of good.
6. This could be due to a negativity bias in characterisation of ethical values concentrating more on negative issues and events rather than positive ones.
7. It means encompassing calls for safety and security and includes beneficence.

## Challenges in principles:

1. Interpretation and conflict: How the same principles are interpreted across various guideline documents. For instance, the need for more datasets to "unbias" AI conflicts with individuals' data security and privacy.
2. Attribution: interpreting which domain, actor, or issue these ethical principles pertain to.
3. Such divergences could undermine attempts to develop a global ethical AI agenda because varied perspectives, for example risk-benefit evaluations, will lead to different results based on whose well-being they are developed for or the actors involved in developing them.
4. For instance, does the European guideline on privacy also apply to China where privacy guidelines target only private companies, and citizens are accustomed to living in a protected society with high trust in their government?
5. Implementation: Whether through government organisations, inter-governmental organisations, industry leaders, individual users or developers, or by harmonising AI agendas across the board.

6. If harmonisation is a goal, then how does one account for moral pluralism and cultural diversity across countries.
7. Environmental Ethics and Sustainability: AI deployment requires massive computational resources, and thus high energy consumption and calls into question the possibility of harnessing the benefits of AI for the entire biosphere.
8. Integrity: means being explicit about best practices and disclosure of errors.
9. Solidarity and Equity: in solving socio-economic challenges like job losses, inequality, and unfair sharing of burdens. For instance, compensating humans whose actions provide data for training AI models.

### **Suggestions for future:**

1. Defining Principles: Admittedly, principles are difficult to translate into practice. However, they still play a crucial role in building awareness and acting as catalysts for building beneficence and a culture of responsibility among AI developers.
2. Internalised norms and values and an effective AI governance strategy will require both—principles encouraging cultural change in the AI community, and explicit rules and regulations buttressing them.
3. A unified regulatory and policy framework on the ethical, economic, and social implications of advances in AI that established clear fiduciary duties towards users.
4. A homogeneous professional standards and moral obligations of what it means to be a “good” AI developer.
5. Strong institutions that ensure ethical conduct on a daily basis.
6. Ad-hoc committees like United States National Artificial Intelligence Advisory Committee (NAIAC) that dispenses advice to the president and various federal officials; the expert group on AI at the Organisation for Economic Co-operation and Development (OECD); the High-Level Expert Group on AI formed by the European Commission; and the Select Committee on AI appointed by the UK Parliament’s House of Lords.
7. Professional and legal accountability mechanisms to redress misbehaviour and ensure that standards are upheld.
8. Private sector involvement like Google, IBM, Intel, Microsoft, and Sony, have released guidelines for developing ethical AI.
9. Civil society including Non-profit organisations and professional associations, such as the Institute of Electrical and Electronics Engineers (IEEE), Internet Society, Open AI, and the World Economic Forum have also issued declarations and recommendations on AI principles and policies.

### **Way forward:**

1. It is time to move towards a principles-led approach to AI and define clear long-term pathways, set explicit professional standards, and build accountability structures that are not only country-specific but also sector- and organisation-specific.
2. Mechanisms should also be set up to license developers of applications with elevated risks, such as facial recognition tools or other systems trained on biometric data. Privacy-preserving techniques, like homomorphic encryption or federated learning, have been developed for the use of data and learning algorithms.