ARTICLE

Ethics in **Artificial Intelligence**

*Yukti Arora & Anju Tandon

Artificial Intelligence (AI) is a branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.

Artificial Intelligence (AI) is not a new concept, it has been in existence since 1950 when researchers first started thinking on the possibility of using machines to simulate human intelligence. However, AI only really picked up in the late 2000s, when several enabling factors grew to a level: like the increased availability of affordable computer power; the rise in the volume and variety of data, and the speed of access to it; and the emergence of new and advanced algorithms which were able to analyze data in a more "intelligent" way. AI was initially used in financial sector for various finance modelling, aircraft industry to design airplane and train pilot, weather forecast, logistics, data mining, medical diagnosis , but almost all industry now have slowly adopted and are using same.

Largely speaking, Artificial intelligence (AI) is the ability of computers or machines to perform tasks that are typically thought of as requiring human judgement or intelligence. The term cognitive technologies are often used interchangeably with AI, with cognitive referring to the human process of thinking. Few examples of such tasks include visual perception, speech recognition, and decision-making and learning under uncertainty.

The artificial intelligence systems are powered by machine learning (ML) – ranging from rules-based to deep learning to artificial neural networks.

There is no single definition of AI. The lack of a consensus on a definition may be explained, at least in part, by the fact that AI is not a technology per se, but rather a collection of techniques that mimic human behavior. AI is an umbrella term that covers a number of technologies and disciplines such as Machine learning – both supervised and unsupervised, Natural language processing, Cognitive systems, deep learning, Image recognition, affective computing and computer vision.

AI can be approached through various technique like evolutionary algorithm, fuzzy logic, expert system based or rule based and machine learning system While ML has enabled significant progress in the field of AI, a combination of techniques, especially ML+ rules-based approach, is typically used to develop AI-based solutions.

Examples of "narrow AI" are Google search, Image recognition software, Siri, Alexa and other personal assistants, Self-driving cars & IBM's Watson.

The "Strong AI," is a machine with general intelligence and is much like a human being, i.e. as shown in movies (Star Trek etc.)

AI research and development has been focussing on following fields:

1. Thought process and reasoning - Thinking humanly & rationally
2. Behaviour - Acting humanly & rationally

Should we keep developing and advancing in the field of AI?

We, humans, have a guiding wisdom sitting in our brain that guides us to operate as per moral principles. Our moral standards are dependent on upbringing and environment, and these are further regulated through ethics, compliances and laws of the land.

AI can drive strategic transformation program, enable development and growth of market, bring operational and cost efficiencies, empower to deliver better and more tailored customer engagement. However, limited availability of the governance, strategy, right quality and quantity of data, insufficient understanding of AI inherent risks, a firm's culture, and regulation can all act as real, and in some cases, perceived barriers to widespread adoption of AI.

All international regulators have taken an active interest in AI, and while they recognize the benefits that AI can bring to financial markets, consumers, and their own work, they are also increasingly mindful of the potential risks and unintended consequences that the use of AI if not governed and driven by ethics may have.

This is particularly relevant given that, over recent years, we have seen many adoptions of AI hit by a significant number of mistreatment of customers and market misconduct. The resulting focus on the fair treatment of customers and market integrity, together with the relatively untried and untested nature of AI, have meant that organization have to be understandably cautious about governance and ethic while considering the adoption of AI solutions. More so, AI is use of algorithms that are capable of learning from data; they can enhance themselves by learning new strategies that have worked well in the past or can themselves write other algorithms.

Thus, if we do not govern them, enabling them through ethics and just make them focus toward result they can go to extreme of doing anything to achieve the result and cause huge risk in area of usage.

Ethics (noun): moral principles that govern a person's behaviour or the conducting of an activity.

Artificial Intelligence can only separate right from wrong based on data labels. AI doesn't have awareness of itself, nor does it have something called “empathy” which is the fundament of ethics. AI is dependent upon how the machine has been coded and how it learns. So, the ethics of AI are dependent upon

a) How the developer has coded and set the bar for right and wrong as well as

b) How it gets trained during usage.

Thus, it is extremely important to define and follow ethics in AI.

In the past five years, private companies, research institutions and public sector organizations have issued principles and guidelines for ethical artificial intelligence (AI). However, there is an ongoing debate about the following:

1. Constitutes of 'ethical AI'
2. Ethical requirements, technical standards and best practices needed for its realization.

So far, global convergence is emerging around five ethical principles - transparency, justice and fairness, non-maleficence, responsibility and privacy

Concerns have been raised that AI may impact jobs, be misused with a malicious intent, elude accountability or lead to unfair assessments leading to systemic risks and unintended negative consequences.

Before developing an AI that can mimic the Human mind, we need to first train them on human value and govern them in an effective manner.

Following are some of area of risk consideration which should be considered while developing effective and ethical working of AI driven solutions:

AI Model & Governance:

Where there is dependence on a continuously evolving dataset that drives AI decisions, it is hard to identify inherent bias in the model and same can only be addressed by regularly monitoring and providing corrective feedback. If this is not effectively governed, they could be Inherent bias in input data which may result in inefficient and/or unfair outcomes. Thus, clear and objective direction of group of personnel including independent review is required to manage this risk. It is essential that data scientists consider Bias element and address same adequately from the start.

Algorithm model success depend on selection of type of algorithm(s) to be applied to a problem, data quality or optimal choice of algorithm parameters. They can be addressed by proper data governance, algorithm evaluation wrt to problem and proving adequate guidance and evaluation mechanism for same

AI solution's ability to produce accurate results depend on regular and adequate feedback, there is need to define policy procedure to address this, monitor the feedback and enable continuous feedback and learning.

Also, users need to be trained to understand complex AI model limitation else there is increased probability that business users may incorrectly interpret AI outputs leading to poor outcomes. Roles, responsibilities and accountabilities must be clearly defined across the AI lifecycle.

Information and Cyber Security:

Organization should govern evaluation and selection of technology component, in case there is Increased dependency on open source components (software packages, programming language, API, etc.) which are no longer supported or updated or freely available by the creator, such selection may introduce additional security vulnerabilities. The risk become more significant since complex algorithms make it harder to understand how the solution reached a decision and therefore may be subject to malicious manipulation, both by humans or other machines.

Change Management:

In AI solution development and implementation, it is often difficult to identify the impact of changes to upstream systems which feed the AI solutions. This may result in unforeseen

consequences for how AI interacts with its environment.

IT Operations:

Significant dependence, in some instances, of AI applications on big data increases the risk posed by existing legacy IT infrastructure, as the latter may not be compatible with AI (e.g. legacy systems unable to process big data).

Regulatory & Compliance:

Increased risk of breaches in relation to data protection legislation (e.g. GDPR), including data subject rights around automated decision making, due to the continuously evolving and opaque nature of some AI solutions.

Difficulty for management to understand, and justify to regulators, how decisions are made in complex AI applications, such as those employing neural networks, which consist of a number of hidden decision-making layers.

There could be Cultural challenge for large scale AI adoption due to actual or perceived regulatory and ethical concerns which vary across geographies.

People Roles and Responsibilities:

Increased risk that roles, responsibilities and accountabilities may not be clearly defined across the AI lifecycle. Lack of continuous engagement, and oversight from stakeholders (compliance, business, IT, coders, etc.) may increase the risk of things going wrong.

Market:

Over-reliance in the market on a relatively small number of large third-party AI vendors increases concentration risk and may have network effects in the event that one of these entities becomes insolvent or suffers a significant operational loss.

Increased systemic risk resulting from herding behavior (i.e. organizations acting identically to other market participants), if algorithms are overly sensitive to certain variables (e.g. stock market prices).

Supplier:

Use of “black box” algorithms may result in a lack of clarity around allocation of liability between vendors, operators and users of AI in the event of damages.

Increased risk of AI third-party providers' failure, especially in the case of new and smaller players, which may not have sufficient governance structures and internal controls in place.

To manage above risk so that essence and utility of AI is not lost and AI remain governed and ethical, it is essential for organization to define and adopt Risk Management Framework (RMF) lifecycle

Details and language will vary from firm to firm, but conceptually an RMF lifecycle comprises four key stages:

1. Identify - Understanding the risk universe by identifying which risks could have a material adverse impact on the organization's business strategy or operations. This stage also involves monitoring the internal/external operating and regulatory environments to identify changes to the inherent risk landscape and ensure the framework remains fit for purpose.
2. Assess - Defining and embedding a risk assessment process to assess the level of risk exposure.
3. Control - Embedding a control framework to mitigate inherent risks to a residual level that is in line with risk appetite.
4. Monitor and report - Designing a methodology for assessing the effectiveness of the control environment, including relevant metrics for measuring effectiveness, tolerance thresholds, and controls testing.

Reporting the status of the residual risk profile, the control environment and remediation programs to the relevant governance fora.

However, as we mentioned earlier, regulators are also increasingly mindful of the potential risks and unintended consequences that the use of AI by regulated firms may have. From a financial stability perspective, potential network and herding effects and cybersecurity are some of the major areas of concern; from a conduct perspective, regulators are mindful of the potential for unfair treatment of customers and mis-selling resulting from inaccurate AI models, the lack of customer understanding about how their data is processed, any increase in financial exclusion, and negative outcomes for vulnerable customers.

As is the case for firms, most of these risks are not new to regulators. The challenge that regulators face with respect to AI, and innovative technologies more generally, is finding the right balance between supporting beneficial innovation and competition and protecting customers, market integrity, and financial stability. The MiFID II rules on algorithmic trading are a case in point.

Imbibing Human values for ethical AI

AI should have defined behaviour guidance duly governed by wisdom. It is essential for AI implementation to include emotional aspects and consider its impact on society, culture, ethos and governance of overall environment. Learning in AI should not only be materialistic aspect of history of transactions and models but also cover aspects of right and wrong as well as its consequences. All views such as socialist and capitalist, should be blended in design of AI for the solution to remain relevant in long run. Rationality which is inbuilt feature of AI

should be strengthened to include important and crucial aspects of human values like compassion, respect, integrity, creativity, tolerance and justice.

Laying governance mechanism, regulating AI are thus important for both the industry and society. It is essential for regulators, industry, and academia researcher to work together and contribute to the cross-border and cross-sectoral debate about the long-term societal and ethical implications arising from widespread adoption of AI. The collective learning should form basis to lay down the appropriate policy response for AI development and Implementation.

In fact, understanding this need certain national and international organizations have already responded to these concerns by appointing expert committees and engaging them to produce reports and guidance documents on AI. Similarly, companies such as Google and SAP have publicly released AI guidelines and principles. Declarations and recommendations have also been issued by professional associations and non-profit organizations such as the Association of Computing Machinery (ACM), Access Now and Amnesty International. All need to be studied together for a well-defined public policy, technology governance and research ethics based AI .

We may conclude by leaving thought of need of ethical guidance, governance and investment of time & effort to draft legally binding regulations to guide and control the advanced research into AI. ■

*__Ms. Yukti Arora__ is *CEO, Aumyaa Consulting Services LLP- a women consulting organisation. By qualification she is a CA, CISA, DISA, with over 21 years of risk and technology consulting across large listed companies and the Big 4 consulting firms. She is also a faculty at the ICAI in the IT assurance, Blockchain and RPA domain.*

*__Ms. Anju Tandon__ *is an experienced professional in data analytics, digital, IT, business & process transformation and risk management solutions; with over 33 years of experience behind her, implementing and consulting on large-scale strategy and solutions.*