



Dr. Anish Agarwal

Global Head of Analytics,
Dr. Reddy's Laboratories

Expert Speak

Of Data, Ethics, and Privacy

As our physical has become increasingly entwined with the digital, the omnipresence of digital technologies has helped make our lives easier. Businesses across the spectrum are leveraging data and analytics powered by machine learning and artificial intelligence to achieve significant competitive advantage, at a progressive rate. Yet, the measures addressing the ethical concerns of the stakeholders have not caught on with the same speed. With technologies getting smarter, it has become essential to define the ethics of the new digital age. How should organisations navigate this new terrain, and reap the benefits of digital progress while respecting the concerns of all stakeholders?

Have you ever experienced this amazing yet scary phenomenon where you suddenly started seeing online advertisements for something you had just thought of? In today's fast-paced, technology-driven world, digital surveillance is not just a dystopian nightmare, but a reality. Every time we do anything online—shop, research, chat, comment, or even scroll through social media—we inevitably leave behind a digital footprint. This can extend offline with locations tracked, logging in the habits you have, the people you meet, and the choices you make. All this surveillance is bound to leave one with an eerie feeling of being constantly followed. This leads us to the question of ethical use of big data.¹

Today, as the nature of work is changing dramatically, the need to uphold and promote ethical practices is greater than ever. Rapid technological advances have impacted how tech platforms and employees across industries capture, store, use, and report data. Data has been a catalyst for the rise of technologies like artificial intelligence (AI) systems, machine learning (ML) algorithms, and the Internet of Things which are capable of using, analysing, and generating large amounts of complex data at a rate most humans can't even comprehend.

However, despite the positive changes and benefits it can have on different sectors and the overall society, if misused, it can compromise privacy, data protection, and ethical standards, often referred to as the 'creep factor' of big data. Data has the potential to shape our future for the better. However, the misuse could lead to negative impact on millions of lives. This is where the need of data ethics stems from. Ethical challenges arise when opinions on what is considered right and wrong diverge; a simple example of this is the use of algorithms that influence consumer choices. With data becoming increasingly crucial in decision-making, it is important to address these concerns promptly.

The first step here is to understand how this data is collected. The answer lies in our everyday, most of which is contained in the digital devices in our hands. With everything connected to the internet, we willingly provide some information given a promise of privacy while some is collected without our knowledge through the cookies we collect in our journey around the internet. Therefore, we can identify two ways in which people share (or are made to share) their personal data and information—consensual and non-consensual.

¹ Big data primarily refers to data sets that are too large or complex to be dealt with by traditional data-processing application software.

In the first case, users are explicitly made aware in a timely manner that their data is being collected and stored and they consent to the same in an informed manner. In the second case, users remain unaware that their data is being collected and stored. The latter is where the real problem begins. Where should organisations draw the line on utilising data of millions of users, for their own benefit?

Big Data, Big Concerns



Given the wide usage of data and data analysis tools across various industries, the need for data ethics is ever-growing. Data ethics refers to the moral principles, values, and guidelines that govern the responsible and ethical handling, use, and management of data. It involves considering the implications, risks, and potential harms associated with the collection, storage, processing, analysis, and sharing of data, particularly in the context of emerging technologies and digital environments. Here are some of the non-negotiable aspects of data ethics that all organisations must incorporate in order to curb wrong use of data:

1. Informed consent: In general, this entails providing unforced and informed permission for actions committed on your personal self, including your physical body, your information, and your belongings. In a data driven world, consent is, without any doubt, crucial and more importantly, ethical. It is imperative for users to know and understand how their data could be used, and only after



completely analysing the pros and cons, should they permit any organisation or individual to access their data. Organisations can obtain informed consent by being transparent about data collection and usage, seeking explicit consent, offering granular options, ensuring consent is freely given, providing clear privacy policies, implementing consent management systems, and maintaining ongoing communication with users. However, as data analytics studies widen their scope of research to discover newer trends and patterns of consumer likes and dislikes, it has diluted the meaning of informed consent.

2. Privacy: The ethics of privacy revolves around the moral principles and considerations regarding the collection, use, and protection of personal information. It involves determining what is right or wrong in terms of respecting individuals' privacy rights and ensuring responsible handling of their data. The concept of privacy can be understood in terms of the right to privacy, the conditions that dictate one's privacy, and the loss or invasion of privacy.



Consider the extensive research that goes behind targeted, online advertising. When browsing the internet, various digital platforms collect data about users' preferences, interests, and online behaviour. This information is then used to create targeted advertisements. However, this practice can lead to a situation where individuals feel their privacy is compromised, as their personal data is utilised without their explicit consent. The ads that follow users across different websites often seem intrusive and give the impression that their online activities are being constantly monitored and tracked. This raises concerns about the balance between personalised advertising and the protection of individuals' privacy in the digital age.

3. Data Ownership: When we speak of ownership in relation to data, we move away from the traditional or legal interpretation of the word as an exclusive right to use, own and sell. In this context, ownership is more related to the redistribution of data, the modification of data, and the ability to benefit from data innovation.



Historically, lawmakers have held that data is not a commodity and therefore cannot be stolen; a belief that offers little protection or compensation to internet users and consumers who provide companies with valuable information for no personal gain.

Contrary to popular belief, the users of social media platforms are not owners of the data that they generate. In fact, this data is basically liquid gold for the social media platforms, who can utilise it for their own research and development or worse, sell it to other companies.

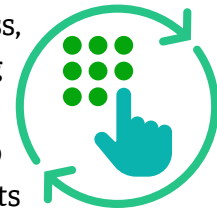
In recent years, the idea of monetising personal data has gained popularity; the ideology aims to restore users' ownership of data and level the market by allowing users to legally sell their data. This is a hotly debated in the area of legislation, with some arguing that commercialising data means losing one's autonomy and freedom.

4. Algorithm errors and objectivity: Algorithms are basically step-by-step instructions that computers follow to solve problems or complete tasks efficiently. However, the objectivity of algorithms can be compromised when they inadvertently reflect the biases, prejudices, or limitations of their human designers or the data used to train them. For instance, resume sorting algorithms have been documented to develop rating biases, rejecting or undermarking resumes based on the examination of the existing workforce of the company.



These biases highlight the importance of continuously evaluating algorithms for potential biases and taking measures to mitigate them, such as incorporating diverse training data and considering fairness and ethics in algorithm design. Addressing algorithmic biases requires a lot of vigilance, diversity in representation, and a commitment to ensuring fairness and objectivity in algorithmic decision-making.

5. Big data divide: The big data divide refers to the unequal access, availability, and utilisation of big data and its benefits among different individuals, organisations, and regions. It exacerbates existing social, economic, and digital inequalities, creating a gap between those who can harness the power of big data for insights and innovation, and those who are left behind, missing out on its potential advantages. Bridging the big data divide requires efforts to promote equitable access, data literacy initiatives, and inclusive policies to ensure that everyone has the opportunity to benefit from the vast potential of big data.



Principles of Data Ethics


Data ethics is of great importance to analysts, data scientists, and technology enthusiasts. However, those who deal with data should know the basic principles of data ethics. While you may not be the person responsible for implementing the tracking code, maintaining the database, or writing and training the machine learning algorithm, an understanding of data ethics can help you understand collection, storage, and the ethical and unethical use of data. This is how you can protect yourself and your customers. So, what are these principles? Let us have a look:





- 1. Transparency:** Includes disclosure of collected data, decisions made using algorithms, and whether the user is dealing with robots or humans. It also means being able to explain how algorithms work and what their results are based on. Transparency also includes how organisations tell people about how their data is used and stored. Sometimes this means that people are allowed to correct or remove information.




2. **Fairness:** Data is not just information, sometimes it gives opinions. Since data is generated by and relates to people, it is crucial to examine the discrepancies in the information collected, the policies framed, and the questions asked pertaining to data. While there is no way in which one can completely eliminate bias from manmade tools, efforts in understanding the causes behind these biases can help reduce, manage, and correct these anomalies to a large extent to improve the decision-making process.


3. **Accuracy:** The data used must be recent and accurate and there should be processes in place, in order to ensure the hygiene of the data. The data must be carefully managed, cleaned, sorted, combined, and shared to maintain its accuracy. When data is taken out of context, it can sometimes become misleading or false. Accuracy therefore depends partly on the reality of the data and partly on its importance and usefulness in relation to what we are trying to do or learn.


4. **Accountability:** Abiding with global laws and regulations is not where responsibilities end. Organisations, in fact, need to be accountable for their use of data. Not only does it play a huge role in gauging the risks associated with use and analysis, but also helps in authenticating the source and quality of the data.


5. **Intent:** In discussing any branch of ethics, intent is important. Before you collect data, ask yourself why you need it, what you get from it, and what changes you can make after analysis. If your intent is to use data in a manner that could cause harm to others, exploit test subjects' weaknesses, or apply it for other malicious purposes, it is unethical to collect their data.



Conclusion

Now more than ever, every business is a data company. It is estimated that by 2025, individuals and businesses worldwide will be generating 463 exabytes of data per day, up from less than 3 exabytes a decade ago.² A few companies have

² Desjardins, Jeff. (2019, April 17). How much data is generated each day? *World Economic* <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/>

systematically considered and begun to address the ethical aspects of data management, which can have far-reaching implications and responsibilities.

Organisations may feel that by employing multiple data analysts, they have fulfilled their data stewardship obligations. However, the truth is that data ethics is everyone's domain, not just the domain of data analysts or legal and compliance teams. Therefore, it is imperative for businesses to train their employees at all levels to understand the use of data and the ethical standards to use it, and also equip them to identify biases and lapses in data collection and interpretation.

***Dr. Anish Agarwal** is a senior leader with over 22 years of experience in the field of data strategy and analytics. Currently the Global Head of Analytics at Dr Reddy's Laboratories, he is also a mentor for NASSCOM DeepTech Club and T-Hub. He serves as an academic advisor to several colleges/ universities across India and has authored more than 15 papers on various subjects related to artificial intelligence and advanced data and analytics. Anish is also a board member of AIM AI forum. In addition to his technical contributions, he's a prominent advocate for data ethics and privacy at the national level, earning him multiple awards and recognitions, such as the AI Young Business Leaders Award, Top 50 Influential AI Leaders in India, and Top 40 under 40 Data Scientists.*