TATA CONSULTANCY SERVICES

Building on belief

Dimensions of AI: Righteous, reliable, and restrictive

Abstract

Enterprises are increasingly leveraging artificial intelligence (AI) across their products and services. However, to retain their competitive advantage and offer differentiating value proposition to clients, they must manage all dimensions of AI solution design, development and deployment to ensure these solutions are safe, secure and sustainable.

For safe, secure, sustainable AI, enterprises must learn to gainfully manage the three dimensions of AI– righteous, reliable, and restrictive– as an integral part of AI and data governance. However, AI governance and data governance have overlapping aspects, and none can be considered in isolation, in the same way as they are part of the larger IT governance. This white paper focuses on the critical aspects necessary to achieve righteous, reliable, and restrictive AI.

Righteous AI

For wider acceptance and sustenance of AI solutions, enterprises need to ensure that the outcome of the AI solution is fair, ethical, and purpose-driven, with the larger goal of achieving societal and environmental wellbeing. Missing out on the righteous aspects could lead to financial loss, loss of reputation, and disruption in business.

Academia, industry, and government are continuously investing in enhancing the righteous aspects of AI through technology advancements, industry ethics, and regulations. Thus, there is no 'one size fits all' approach to being righteous; it depends on the context of the business process. Defining and practicing what is righteous is both art and science, but by following four key dimensions, what we call the 4Bs, enterprises can well manage this aspect of the AI solution.

Baseline the existing approach
Businesses have been operating on human skills, based on norms for what is righteous and what isn't. As enterprises adopt AI solutions to augment human skills to a scale that is beyond human capacity,baselining the existing righteous aspects from the perspectives of customers, industry, partners, regulators, and government will become critical.

Build awareness
AI solutions work on the principle of continuous learning, wherein human involvement is critical so that the solutions can be trained and tweaked to perform as desired. In such a scenario, it is important to exercise care that AI solutions do not end up performing actions that are against the enterprise's acceptable righteous guidelines. Thus, awareness-building around the righteous aspects of the solutions and how to manage them are important.

Blend theory and practice
As the evolving righteous aspect of AI solution is both an art and a science, enterprises need to

blend the theory and practical aspects. A balance must be struck between the theoretical systemic approach and the real-life business situations for an enterprise, based on its operating environment and ecosystem (taking into account customers, partners, vendors, regulators, national/regional/international laws).

Business process empowerment
While the enterprise should have an overarching framework for managing the righteous aspects of the AI solutions, each business process should analyze what defines righteous for that process. More importantly, the definition must be granular at each step of the business process so that what is righteous is reflected with integrity in the whole process.

Reliable AI

To make AI solutions reliable, from the perspective of both functional and non-functional performance, enterprises would need to manage data bias, data poisoning, and model stealing.

Data bias
This takes place when training data does not gainfully represent the problem statement for creating the AI solution because of incomplete, inaccurate, influenced (or manipulated), and imbalanced data. The negative impacts of data bias can be significantly reduced, if not eliminated, by following the 4Cs.

- Clarity: Having clarity about what is expected from the AI solution and what data is needed for the same
- Correctness: Ensuring correctness of the data through proven data analysis methods
- Checks and controls: Building them in to validate the AI solution at each lifecycle stage
- Configure and reconfigure: Training and tweaking the model periodically or when there is a discrepancy

Data poisoning
It adversely impacts the training of the AI solution, causing it to behave maliciously. Data poisoning can happen irrespective of the knowledge of the solution, in two ways:

- White box, when the attacker has the know-how of the AI solution
- Black box, when the attacker does not have the know-how

The source of data poisoning takes place through:

- Injection: The solution malfunctions due to injection of corrupt data, inclusive of labels of data.
- Infiltration: The attacker tweaks the solution to get the output they want.

The impact of data poisoning can be significantly reduced, if not eliminated, by, first, focusing on data cleaning and data pre-processing for outlier and irregularity detection, and second, by assessing the functioning of the AI solution by making it process new data sets.

Model stealing
Model stealing, so that unethical acts can be performed for undue advantages, typically happens in the following ways:

- Executing the AI solution with different data sets and then analyzing the results

- Carrying out an exercise aimed at finding possible different weights assigned to the independent variables of the AI model
- Extrapolating the scope of the solution for solving the specific business problem

Model stealing can be significantly reduced, if not eliminated, by deploying the 4R technique:

- Restraining the access of the AI solution to only those who need it
- Restricting the usage of the solution only for the cause intended for and agreed upon
- Revealing limited information about how the solution works as needed to limited people
- Reviewing and curbing the provision of complementary information about the solution's purpose and functioning

Restrictive AI

In the AI world, while technology advances demystify the data, both structured and unstructured, AI needs to learn from the data as well, making the management of data privacy crucial to preventing any undesirable consequences.

Data privacy is all about ensuring that data is rightly handled through the cradle-to-grave life cycle of the data, that is sourcing, storage, processing, usage, sharing, archival, and deletion. The right handling of data should not be focused just on managing regulatory compliance, but should also be driven by self-discipline. With adoption and adaption of the 6Ds, enterprises can effectively manage the demands of data privacy, as a part of the overall data governance framework of the enterprise. This would ensure that the focus on data privacy is proactive and not reactive, and that the integrity and confidentiality of the data are not compromised.

Data sourcing
In a data-dominated world, enterprises need to leverage all the available data, comprising data generated within the enterprise, data procured from outside, data collected from the ecosystem stakeholders, and open data. Irrespective of the mechanism of data sourcing, enterprises must ensure that they have a consistent policy to manage the privacy of the data. While making the data available to AI solutions, they should ensure that the AI solutions should be focused on what is expected from them rather than managing the nuances around the data sources.

Data usage
The AI solution should only use the minimal data necessary to achieve its intended objective, and what is legally permissible to use. It is very important that anonymized data be used, and not the associated PII (personally identifiable information). Wherever possible, enterprises should evaluate usage of GAN (Generative Adversarial Network) for generating the synthetic data. In all cases, the AI solution should have complete transparency around what data it is using, and should demonstrate that it is respecting data privacy rules.

Data processing
AI solutions typically process huge volumes of data and that too from a number of sources, with multiple classifications. The data is typically classified as public (available to all), internal (available to all employees), confidential (typically individual or business function-specific data) and restricted (available to specific business functions and roles). The AI solution should not be accessing the data it is not permitted to access. Further, as AI solutions process the data multiple times and in multiple iterations, it should be ensured that the data is not exposed in any form in the process, leading to privacy compromise.

Data storage and archiving
AI solutions should strictly follow the policies defined around data privacy during storage, be it persistent storage or data in transit, or data transferred for end-user application usage. Security considerations around data encryption and access control should be respected by the AI solutions as long as that data is in its control.

Data sharing
Sharing becomes critical during the evolution of the AI solution, especially the purpose of sharing, with whom the data is getting shared, why it is getting shared, for how long the data is to be shared, what exactly is getting shared, and is data shared free or at a price? Similar considerations are applicable while sharing the solution output.

Data deletion
In the context of AI, data deletion, or purging enterprises must take care of both output data and in-process or intermediate data. The output data also has related log data, which is typically associated with the explainability of the AI solution, and thus if output data is deleted, the corresponding logs will have to be deleted. The AI solution needs to produce intermediate data before output data – the retention policy of this data is driven by business needs. In any case, one must ensure that the intermediate data gets deleted along with the output data.

Conclusion

AI continues to provide enterprises with compelling solutions to problems so far considered to be beyond human intelligence. However, a comprehensive approach is required to meet the challenges around AI solutions in the righteous, reliable, and restrictive dimensions. This approach will have to continuously evolve as enterprises periodically review the righteous, reliable, and restrictive dimensions to secure more benefits from their AI solutions.

About the author

Mahesh Kshirsagar
CTO – Analytics & Insights, TCS

Mahesh is responsible for shaping innovative and purpose-led solutions that accelerate the growth and transformation agenda of enterprises. Mahesh is an engineering graduate, who started his career in TCS in 1990, and has IT expertise of 30+ years, spanning technology domains, industry verticals, and software processes. Over the years, he has incubated several high-impact, business-aligned IT solutions and services having high revenue potential, focusing on thought leadership and innovation to enable growth and transformation. His IT expertise stems from his employment experience in system integrator companies, end-user organizations, IT products, and BPO organizations.

Conceptualizing, architecting, and delivering state-of-the-art business IT solutions is his strength. He has applied for more than 30 patents for his solutions, of which around 10 have been granted. His solutions have also won several prestigious industry awards.

Awards and accolades

TOP 3
IT SERVICES
BRAND

FASTEST GROWING
IT SERVICES BRAND
FOR THE DECADE
2010 - 2020

Contact

Visit the Analytics and Insights page on https://www.tcs.com

Email: BusinessAndTechnologyServices.Marketing@TCS.COM

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is a purpose-led transformation partner to many of the world's largest businesses. For more than 50 years, it has been collaborating with clients and communities to build a greater future through innovation and collective knowledge. TCS offers an integrated portfolio of cognitive powered business, technology, and engineering services and solutions. The company's 500,000 consultants in 46 countries help empower individuals, enterprises, and societies to build on belief.

Visit **www.tcs.com** and follow TCS news **@TCS_News**.

Corporate Marketing | Design Services | M | 08 | 21