LTI
Let's Solve

Embedding Privacy into Big Data Analytics

By Swati Koul

Never has there been a greater sense of urgency to strengthen data privacy and its protection. Technology is enabling companies today to use Big Data, Artificial Intelligence, and Machine Learning to conduct business as usual. As data increases, traditional data sets are being stretched to include social media data, browser logs, and text analytics to acutely gauge customer mindset. This goes to the reason why big data analytics has undergone a significant overhaul in the last few decades. What started as a descriptive concept soon became predictive and now shows advanced prescriptive symptoms. This way, companies have gone from understanding consumer behavior to influencing it and now controlling it.

Big Data analytics, AI, and ML provide numerical evidence on all aspects of consumer lives – health, personal preferences, buying habits, loyalty, etc. thus eroding personal privacy without restraint. They highlight the information asymmetries over common economic actors, thus shaping the economic performance of businesses & countries.

The aggressive use of algorithms, opacity of processing, proclivity to collect 'all the data,' repurposing of data, and use of new types of data all have potential implications for data protection. Since these self-improving algorithms use data as fuel, they inadvertently challenge purpose limitation, data minimization, data retention, transparency, and the idea of consent. In short, everything that the GDPR aims to preserve. How can then companies apply AI and still abide by the GDPR?

There is no straightforward data-processing approach since analytics tends sometimes to go wayward and use data in unexpected ways. However, the companies that handle personal data need to factor in the effects of their processing on the individuals as part of assessing fairness and communicate the same to maintain transparency.

Instruments such as privacy notices that explain the purposes for collecting data can aid the GDPR compliance. Companies can also rely on people's unambiguous, clearly-stated consent given that people understand how this data will be used. They can adopt 'just in time' notifications or graduated consent, where people agree to varied uses of their data throughout their relationship with a company, rather than be restricted by a binary choice at the start. Additionally, they can use analytics when processing is 'necessary' to serve legitimate interests like fraud prevention or IT security.

Since Big Data analytics tend to repurpose data, companies need to carry out a compatibility assessment of processing purposes to abide by the GDPR. For example, if the repurposing is to detect trends or correlations, it advocates a clear functional separation between analytics operations. However, if it is done to predict consumer behavior and make decisions affecting them, an informed and unambiguous 'opt-in' consent is required to ensure further use is valid.

The GDPR also emphasizes data minimization and retention. By contrast, big data analytics tends to favor excessive data collection, notwithstanding the relevance of the volume collected. Though the 'right to be forgotten' lawfully empowers individuals to better control their data, companies, on their part, need to be able to articulate at the outset why they need to collect and process specific datasets. They must be clear about what they expect to learn or be able to do by processing that data, and thus prove that the data is relevant and adequate. To enforce appropriate retention schedules, they must adopt structured data governance and delete long runs of historical data beyond the period required for regular business purposes.

In place of all the complications that arise from using analytics, AI, and ML, techniques such as anonymization, can help companies meet their data protection obligations. A truly anonymized dataset renders it impossible to identify an individual from the said data or its variations.

Conclusion

If companies don't need to use data that identifies individuals, they can use analytics on anonymized data to gain insights, identify patterns, draw similarities, or detect anomalies without compromising the sanctity of the GDPR. They must show that the risks of re-identification have been evaluated and equivalent solutions adopted. Similarly, embedding privacy by design & default into analytics by employing techniques like differential privacy, that involves injecting noise into the answers of dataset queries, can prevent data misuse and enforce data segregation while retaining the value of the solutions. Despite the vast expanse of the GDPR, the benefits of big data need not come with a trade-off between utility and privacy. It is time that companies conceptually shift from the notion of 'big data versus privacy' to 'big data with privacy.

Swati Koul

Lead- GDPR Compliance Program, LTI

In her 10 years of experience with LTI, Swati has led and delivered large scale cross-functional projects across sectors such as Manufacturing, Media & Entertainment, Consumer Goods, and Banking for Fortune 500 clients. Her areas of expertise include Business Development, Project & Portfolio Management, and Technical Consulting. In her current role, Swati leads the GDPR Compliance Program at LTI, where she has been responsible for developing the endways service portfolio, which has been positioned in the Leaders quadrant of many analyst reports.