

How regulatory pressure is reshaping big data as we know

Dec 28, 2018

Jelani Harper

It should come as no surprise to anyone monitoring the burgeoning big data ecosystem that the increasing quantity of data being generated, the majority of which is unstructured—combined with the growing number of external data sources and quotidian nature of data breaches—has led to today’s hyper-sensitive regulatory environment.

It was only natural that the ability to automate the processing, production, analysis, and management of big data via cognitive computing dominated the epoch in which real-time transactions (in the cloud, via mobile technologies and ecommerce) became the norm. The rapid dissemination of personally identifiable information (PII), the expansion of its definitions, and the inherent incongruities between regulations were similarly logical conclusions of the same vector in which automation and decision-support were esteemed.

But when these same big data developments led to issues of interpretability and “explainability,” and when people or intelligent systems simply relied on quantifiable algorithmic outputs with limited understanding of their biases or the reasons behind them, intervention—in the form of regulatory mandates and penalties—also quite naturally arose.

Some are international in scope and jurisdiction, such as the recently implemented [General Data Protection Regulation](#) (GDPR). Others such as the [California Consumer Privacy Act](#) have yet to go into effect and companies are still working to understand their implications. Many are rigidly exacting in focus, such as the assortment of regulations pertaining to vertical industries, including finance and healthcare.

Nearly all, however, are designed to protect and support the consumer, the data citizens whose lives and information are used to dictate, and in turn are dictated by, the regularization of big data.

The effect is a reconfiguration of how big data is perceived, which will no longer be confined to popular conceptions of artificial intelligence (AI), the Internet of Things, and blockchain but also include the practicality of their use, such as the following:

◆ **Interpretability and explainability:** Statistical cognitive computing models generate numerical outputs informing decisions for everything from personal finance to healthcare treatment. Correctly interpreting those numbers and understanding their derivation are both increasingly difficult due to complex neural networks and deep learning. The issue for data scientists, regulators, and

consumers is “these kind of black box solutions where you’re not sure exactly how the data’s being implemented,” noted Brandon Purcell, **Forrester Research** principal analyst.

◆ **PII and privacy:** Regulatory entities vary by country, industry, and state, exponentially multiplying the data classifications (such as PII) needed to satisfy them. There are “data breach notification requirements across now all 50 U.S. states,” observed Barbara Lawler, chief privacy and data ethics officer at **Looker**. Implicit in these regulations is a de facto shifting of ownership from the organizations that retain the data to the consumers whose data they possess, prioritizing privacy over accumulation.

◆ **Automation management:** Monitoring, validating, and governing cognitive statistical models that automate big data processes reinforce trust and data quality. This necessitates that organizations “look at the performance of those models and have the same governance on the release and changing of them as they have with the data itself,” said David Downing, EVP of product management and product marketing at **ASG**.

◆ **Risk:** The possibility of data loss, data quality, and user errors as well as concerns about regulatory compliance all contribute to big data risk, making it clear that “keeping data is not always a good idea,” acknowledged Dennis Chepurnov, marketing principal at **Hyland**. Solutions to address these concerns include encryption and layered cloud approaches.

Each of these factors redefines big data’s meaning and implementation according to regulations, reflecting the emergent realization that, “We really do need to think about whether we’re creating these models in a way that treats people accurately,” Purcell noted.

<http://www.kmworld.com/Articles/Editorial/Features/How-regulatory-pressure-is-reshaping-big-data-as-we-know-it-129189.aspx>