Big Data, Cognitive Bias and Data Quality's New Frontiers

Make no mistake: Big data is a disruptive, transformative and game-changing force for organizations of all sizes, industries and geographies. The dramatic explosion of inputs and outputs from devices, applications and data sources has accelerated not only in variety and volume, but also especially in velocity. Not only is data being captured and catalogued faster than ever, but it also must be turned into actionable insight faster so organizations can make smarter decisions that give them a competitive edge.

But, according to expert speakers on a recent web seminar, taking advantage of all that big data has to offer means understanding, confronting and overcoming issues such as cognitive bias in getting to the heart of the matter when it comes to using all that data: Trust. Without full and unhesitating trust of the quality, veracity and fit of the data thrown into the hopper of big data engines and analytics systems, organizations will never know if they are making the right decisions at the right time for the right reasons.

TRILLIUM SOFTWARE®

Taking advantage of all that big data has to offer means understanding, confronting and overcoming issues.

"Business decisions are driven by trust in the data, in the information produced from that data, and the underlying models built out of the data that support those decisions," said Harald Smith, director of product management at Trillium Software, which sponsored the webinar, entitled "Big Data, Cognitive Bias and Data Quality's New Frontiers." And trust is just one of the significant challenges in helping organizations really put big data to work, he added.

- Lack of trust in data needed to make rapid, accurate decisions that grow the business.

- An inability to access or understand data, resulting in allocating more time on data preparation.

- A need to facilitate business collaboration and achieve data transparency.

- Overcoming resource limitations in order to empower the business to do more, faster.

One of the keys to understanding how to overcome these challenges is recognizing the existence of cognitive bias, according to Smith. After all, data and data sets are creations of human decision, and cognitive bias certainly affects the basis of many decisions. "The context to understanding (the information) is how it was created, and there may be particular things I'm thinking about or I'm expecting in this particular case," he said. Faced with uncertainty about key facets of the data, the user applies their "own cognitive bias, and out of all these factors, I'm left with a sort of 'well, maybe' [in the decision-making]. Maybe this is what I need to decide at this point in time, but maybe not. I'm not really sure." At this point the user falls back on prior experience, not data.

Assumptions and biases about the validity, veracity and applicability of different factors often have substantial impact on what we think we are learning from big data solutions. Context is essential because there are subtle differences in the meaning of a word or phrase, or data quality can be skewed by decisions to exclude certain segments of a population or data set.

"Big data produces value when it is trusted."

- Harald Smith

And the impact of bias can extend beyond day-to-day decision-making. There are important legal, regulatory and governance statutes where the result of even unintentional cognitive bias can have big financial, customer trust and brand reputation impact.

Ultimately, achieving trust in big data efforts boils down to four issues, according to Trillium's Smith:

- **KNOW YOUR GOAL.** First, organizations must have a deep and common understanding across departments and groups about what they are trying to achieve, what questions they must ask about the data to ensure trust, and what information is necessary in order to make the decision.

- **UNDERSTAND YOUR DATA.** It's essential to have upfront conversation with all key stakeholders to weed out or have a better grasp on assumptions shaping the problem and the decisions. Be sure to keep in mind issues such as the completeness of the data (including its coverage of the relevant data population), the integrity and uniqueness of the data, its consistency and timeliness, and factors such as provenance and user ratings.

- **DETERMINE IF THE DATA IS FIT FOR THE INTENDED PURPOSE.** Not all data is equally relevant—or even relevant at all—for the goal in mind. Everything starts with the business requirements, which in turns shapes data requirements and key metrics.

- **DOCUMENT AND COMMUNICATE YOUR FINDINGS.** Recording the findings improves collaboration and ensuring decisions based on the data align with the intended goals. Open communications also ensures that all key hypotheses are discussed, vetted and evaluated along a common set of requirements, metrics and priorities, and has the ultimate acceptance by stakeholders, influencers and approvers across the organization.

"Big data produces value when it is trusted," said Smith. By knowing the organizational goals, understanding the data, identifying its fitness for the intended purpose and documenting the findings for essential communications, IT and business stakeholders alike can dramatically improve their chances at achieving the essential trust in the ever-growing and diversifying volume of data used for decisions.