World Futures Review
1–10

DOI: 10.1177/1946756720905639
journals.sagepub.com/home/wfr
SAGE

Mine the Gap: Augmenting Foresight Methodologies with Data Analytics

Anne Boysen[1]

Abstract

The explosion of Big Data and analytic tools in recent years has brought new opportunities to the field of foresight. Big Data and improved analytics capabilities can expand the knowledge base and act as a corrective to our cognitive biases. Moreover, several data mining and machine learning techniques that increase performance for businesses can be applied in foresight to help researchers discover patterns that may be early signals of change and correct our misperception of patterns where they don't exist. This article discusses the opportunities and limitations of various data mining and machine learning techniques in foresight.

Keywords

foresight, future studies, data analytics, machine learning, artificial intelligence, methodology

> Without data you're just another person with an opinion.
>
> —W. Edwards Deming

The explosion of Big Data and analytic tools in recent years has brought new opportunities to the field of Foresight. Information gathering and processing that once took weeks and months can now be accomplished in much shorter time and with fewer resources. With the increased data access and analytics capabilities comes not only speed and accuracy, but also better opportunities to study data directly without interpreting intermediaries, such as journalists, publishers, and research institutions. Walls that once existed between foresight professionals and raw data are crumbling, or at least becoming more penetrable, as both the access to and the analytic capabilities of Big Data become ever more available.

While text mining tools that automate environmental scanning are gaining more attention, little has been written about applying statistics, data mining, and machine learning techniques to discover novel patterns directly in primary data. While this article discusses applications of analytics of both primary and secondary sources, it will specifically make a case for the former.

Can Big Data and Analytics Help Fight Cognitive Bias?

The centrality of data and empirical deduction has waxed and waned in philosophy and academic research. More than half a century ago Karl Popper popularized the Hypothetico-Deductive method which has become widely adopted in social sciences as a means to fight positivistic assumptions or theories without corroborating evidence. In his seminal work *The Logic of Scientific Discovery*, Popper

[1]University of Houston, TX, USA

Corresponding Author:
Anne Boysen, University of Houston, 4800 Calhoun Rd, Houston, TX 77004, USA.
Email: anneboysen@gmail.com

(2002, 18–24) warns that "it must be possible for an empirical scientific system to be refuted by experience" and that "a subjective experience, or a feeling of conviction, can never justify a scientific statement." Later schools of thought have posited that there were more solid barriers between the researcher and the objective truth than merely collecting and analyzing data. Social privilege, language, and culture influence the interpretation of reality and how we perceive and emphasize the data we have access to.

Discourses in Foresight have followed a similar trajectory, alternating between empirical deduction and forecasting to more critical approaches that question metanarratives and values (Inayatullah 2009). After all, data from the future on which to make falsifying statements do not exist. The various courses of events a futurist must consider will often depart from current reality in both kind and magnitude, making empirical data from the present less useful when envisioning alternative futures. Questioning common default assumptions is often seen as a more viable approach than predicting the future (Dator 2009).

If our default assumptions color our selection of data sources, it might seem as if data analytic approaches fail to correct cognitive bias. After all we can decide to include some sources and not others. But while the selection of data sources is left to human judgment, we cannot necessarily infer causal relationships *within* the data. Unlike authored information often curated by personalized algorithms, raw data lack narrative and incentives to focus on some elements at the expense of others. It is reasonable to assume that we are less likely to fall for selection bias in our overall analysis if we can better prevent personal interests to influence our analysis.

In a time when "post truth" has entered our dictionary ("Word of the Year" 2016) and anti-scientific factions weaponize warped readings of post-modernism to advance relativistic agendas, efforts to rebuild intersubjective consensus around verifiable facts might be more important than ever (Kuntz 2012; London School of Economics and Political Science 2017). The subjective pitfalls Popper warned about in the mid-twentieth century are still valid, and a strong foundation in data can help us build a solid base of empirical evidence that helps our perception of reality with empirically deduced logic rather than subjectively induced assumptions ("Confirmation Bias and the Power of Disconfirming Evidence", 2017). An empirical approach could be our best defense against biases amplified by social influence and echo chambers (Mounk 2018).

Discovering interconnection, correlations, and causal effects in the systems we want to understand helps us understand current dynamics or locate early warnings of change. If we want to contemplate future state t_2 or t_3, we should first obtain fine-tuned information about the current state t_0. While comprehensive data analysis may have fewer direct applications for fleshing out complete scenarios, our assumptions around data, casual connections, and change should be informed by rigorous data gathering and analysis.[1] If we employ analytics customized to the type of insights we want to unearth, we reduce the chances that our research is biased or misaligned with our research objectives. By doing our own primary data analysis, we can ensure a more future-oriented problem focus and also reduce the chance of sponsor bias, which can be hard to detect when we rely on secondary reports and desk research (Sarniak, 2015).

Of course, a mere focus on data is not enough to remove cognitive bias from research since the interpretation of objective data is bound by the epistemologies of the researcher. Thus, an emphasis on comprehensive data approaches should not be used to trivialize critical discourses around ontology and paradigms. For example, Inayatullah does not argue that critical analysis of the metaphors that surround data should in any case sacrifice the data collection effort. In fact, he suggests that the Causal Layered Analysis (CLA) framework should not be used at the expense of data orientation (Inayatullah 2014). It might therefore be more appropriate to view data centric and more post-structural approaches as synergistic rather than two epistemologies where one precludes the other. Rather data should be seen in the context of meanings, and meanings in the context of available data. Since machine-analyzed data are intrinsically more value-neutral and comprehensive, we should expect a more robust foundation for the critical analysis it enables. A student in my Data Analytics class in the Foresight program at the University of Houston obtained granular knowledge around what makes a person likely to consider opportunities in the emerging gig economy. Building a model that considered a complex combination of demographic and attitudinal variables, she was able to sketch out the combination of traits of people who will be more or less likely to thrive in the gig economy. This knowledge would be useful to scenario exercises focused on the future of work, job automation, or industrial reorganization. In its absence, our scenarios might be informed by anecdotal case examples or loosely applicable secondary findings.

Environmental Scanning through Direct Observation in Big Data

Futurists strive to obtain somewhat similar attributes as data scientists in their information quests. Keeping an eye on external forces of sociocultural, technological, economic, environmental, or political nature, futurists often adopt the STEEP-framework which allows for the detection of early change in the macro-environment (Bishop and Hines 2012). And while futurists consider criteria such as credibility, novelty, likelihood, impact, and relevance of their sources (Bishop 2009), the four Vs of Big Data—volume, velocity, veracity, and variety—are essential attributes of machine learning pipelines (Dea 2015). The main difference between futurists' and data scientists' data gathering methods is that while the former method is inductive, limited, and prone to availability bias, a data scientist can amass and analyze humanly impossible amounts of data with less interference of subjective perceptions.

Essential to the environmental scanning process is identifying inputs which through a process of analysis, interpretation, and prospection render an outcome that can be used in strategy formation (Voros 2003). Early in the information gathering process a foresight practitioner must identify early signals and trends. Conway (2006) distinguishes between trend spotting and trend analysis, pointing out that analysis considers the existing themes and patterns in society. Merriam-Webster defines "trend" as a prevailing tendency or inclination, a general movement. While a trend reflects current events that, based on the frequency of its mention in blogs, twitter, news stories, and so on, can be gauged as either increasing or decreasing, an emerging issue is a latent issue that has not yet reached mainstream attention. Richard Lum writes that a trend is a

> historical change *up until the present*, then an emerging issue is a possible new technology, a potential public policy issue, or a new concept or idea that, while perhaps fringe thinking today, could mature and develop into a critical mainstream issue in the future or become a major trend in its own right. (Lum 2016)

Dator (2018, 7) describes emerging issues as the "far left tail of the 'S' curve of growth, barely visible, and just beginning to pop into view." Hiltunen (2010, 15) writes that weak signals are information about emerging issues and potential future changes. Since weak signals and emerging issues often signify a potential disruption rather than a continuation of a trend, we cannot necessarily rely on continuous data from the past.

Unlike trends, which we can quantify via observable frequency metrics, weak signals and emerging issues are more vulnerable to selection bias on part of the researcher as well as the medium reporting the signal. These biases are even more difficult to harness as search engines and news aggregators increasingly personalize our information. Foresight researchers must confront not only their own cognitive biases, but also the algorithms that are optimized to exploit these biases. Paradoxically, the more immersed in data and algorithms we become, the more critical is it that we skillfully deploy carefully chosen algorithms to navigate our information journey. In a comprehensive literature review, Mühlroth and Grottke (2018) analyzed fifty foresight articles that used a combination of expert and computer-driven scanning for weak signals. The authors found that queries that were led by experts were more prone to human bias than those that were fully computer-driven. While recognizing that human experts are more effective in the later stages such as in strategic decision making and implementation, data mining and automated approaches in the early stages of the corporate foresight process search strategies are required to greatly reduce the human actor bias.

While trend analysis and forecasting techniques have long traditions in foresight, the study of data mining techniques for weak signal identification have shorter traditions. Most data mining approaches currently used in corporate foresight seem to revolve mostly around searching, scraping, and summarizing written text. The data entities are therefore the coded ideas rather than the phenomena themselves. Hiltunen (2008) found that most futurists find weak signals from sources such as politicians, domain experts, journalists and other futurists and direct observation of ordinary people. With the exception of the last category, these are sources that merely express the ideas of change rather than the change itself. Without dismissing the importance of these reports of change, we risk adopting biases that are embedded into the mediation of these reports. Big Data analytics on the other hand makes direct observation of early signals more feasible. One example is when businesses learn about new customer behaviors from data mining their sales data (Attewell and Monaghan 2015; Linoff and Berry 2011). In this context, scanning for weak signals might have more in common with precursor analysis where subcultures, jurisdictions, or early adopters may indicate later change in larger, more slow-moving systems (Molitor 2017). By discovering smaller changes in limited datasets, it might be possible to infer or deduce more impactful changes.

User-Generated Content and Unstructured Data

Which types of data mining methods we chose largely depends on what type of data we use. We can distinguish different type of data based on measurement level and whether the data are structured or unstructured. While structured data analysis has decades-long roots in business intelligence, many businesses still struggle to make sense of their unstructured data. With a ratio between unstructured to structured data of 8:2 (Das and Mohan Kumar 2013), this leaves a lot of potential insights in the dark. Although jurisdictions surrounding the data's origin restrict access and use, unstructured data are more often publicly available than structured and preprocessed datasets. There seems to be an inverse relationship between data access and data cleanliness. While online user behavior such as URL visits, click through rates, and web cursor activity easily convert into machine readable formats, access to such insight is regulated and typically restricted to internet service providers, search engines, and website owners. However, vast sources of user data are available in the form of blogs, discussion boards, social media (Kayser and Blind 2017), and even aggregate-level search history (Stephens-Davidowitz 2017). This type of data typically requires preprocessing and is analyzed using natural language processing (NLP) and natural language understanding techniques.

Interestingly, while AI (artificial intelligence)-driven foresight aggregators such as Shaping Tomorrow's Athena make sophisticated use of unstructured data, foresight professionals seem less inclined to make use of existing datasets data and resources found in their clients' or employers' relational databases.[2] Straightforward statistical or machine learning models can be applied to public and proprietary datasets for valuable insights. Why this is not more common could be due to the common problem with data silos and corporate cultures where foresight professionals rarely cross paths with analytics departments. For subcontracting consultant there may also be confidentiality boundaries that prevent the foresight professional's access to proprietary data. As outside foresight professionals, it might be useful to discuss data policies with clients before embarking on new projects. Futurists working internally in organizations should make connections across departments to gain access to data and insights used for other purposes than foresight.

Unsupervised Learning Detects Signals—and Noise

While text mining and NLP-methods that parse and analyze volumes of text will continue to be a core element of AI-driven foresight tools, machine learning methods offer untapped potential for pattern recognition. Since these methods use classification and clustering techniques, the foresight professional will gain more robust scientific support for their findings. By quantifying potentially interesting observations or search for new patterns in existing datasets, we can find weak signals that indicate change. In other words, machine learning fills the gap between text-based data mining and traditional statistical methods.

Let's for example consider shifting consumer behavior. We might have a hunch that two customer groups, which otherwise have little else in common, suddenly show similar behaviors or attitudes. This could indicate a shift in underlying values in these two groups and affect how we should view stakeholders in our alternative futures. Both structured and unstructured data can help us discern interesting signals. Using text mining to scan the behaviors of these target groups in social media, we might find word frequencies, combinations, and sentiments around relevant mentions. If we have access to demographic data, geolocation, or other metadata we might be able to find approximate correlations harboring the weak signal of a new emerging issue. However, when these data are not accessible or require too much preparation to be practical, we can gain new insights by pulling together various existing databases or sometimes survey data to explore for patterns that may spawn a new trend. Machine learning techniques such as distance-based optimization or cluster models allow us to view signals in multivariate vector spaces that are too complex for the human mind to handle intuitively. Hence instead of using human heuristics to infer relationships, we can view signals contextually using statistically deduced models. Not only does this add scientific support to the observed pattern, it also helps us avoid the temptation of inferring patterns where none exists, a phenomenon called *apophenia*.

Companies sometimes accidentally discover weak signals in their existing data while conducting more conventional analytics exercises (Harryson et al. 2014). Given the irregular nature of weak signals, identification lends itself to computational anomaly detection. In larger datasets, irregular data that could be signals of change are usually paired with other attributes, which makes it possible to learn more about the wider context in which the irregularities occur. By applying rules to detect associations between attributes, we can study emerging new trends by proxy of these more frequently occurring associated attributes.

Supervised Methods and Predictive Analytics

Often when we think of AI we think of predictive analytics. "Prediction" is sometimes seen as a contentious word among futurists because it suggests that the future can be predicted, an assumption many futurists reject. Unfortunately, the term is not often explained precisely in mainstream media. In machine learning, predictive analytics is mostly understood as a classification model that can predict certain outcomes on a target variable based on the correlations it finds in labeled training data. The meaning of the word "prediction" differs both in kind and in magnitude from the way the word is commonly used in foresight. Few data scientists would claim predictive powers for large complex systems, especially where the interconnections are unknown. Moreover, predictions are assigned specific probability values, and few would argue that their models have absolute predictive power. Fit statistics are performed to alert the researcher if they might have run into an overfit model rather than a model which perfectly fits external reality (Attewell and Monaghan 2015, 32–35). Embedded into these models are techniques that quantitatively find levels of accuracy, precision, and recall which tell us how good the model is at identifying occurrences and also predicting their likelihood (Koehrsen 2018). Nor are "black swans" likely to be discovered by predictive analytics (Taleb 2007) since these models look for patterns in past data.[3] It is also important to keep in mind that correlation does not always indicate causation (Pearl and Mackenzie 2018). Predictive analytics informs us *what* is likely to happen, but not necessarily *why*.

Predictive models have proven effective in identifying new customers, victims of crime, patients with certain health risks, and other use cases where we need to do individual predictions for each observation. Since actionable microlevel insights is a common goal when using predictive models, they might be found to be of less direct relevance to futurists searching for change in larger systems.

However, predictive analytics may have some direct application for pattern recognition and some indirect applications which will be addressed later. Decision trees for example render informative visualizations that can build a deep understanding of hidden connections and illustrate connections intuitively to a larger audience. With this application of predictive analytics, we consider not only the individual outcomes, but also the connectivity between each data point. Decision trees in this context can be used as an alternative to other multivariate approaches. And while parametric multivariate approaches that are used to find central tendencies usually require numeric interval data, decisions trees can handle categorical data (Gunluk et al. 2019).

Artificial Neural Networks (ANNs) and New Types of Data

The most sophisticated form of predictive models are neural networks, especially what we know as deep learning. ANNs mimic the highly complex nonlinear structure of neurons in the human brain. An input layer, such as an image, feeds information that triggers the firing of "neurons" connected in various hidden layers until the output more or less resembles the original (Boysen 2019). ANNs are a type of predictive analytics capable of handling data that are of a different type and level of resolution than the aforementioned methods. While lower level data analytic methods limit the datapoints to those that are perceptible to human reasoning, ANNs can handle data rendered in different formats. ANNs can analyze elements that are not necessarily of direct relevance to futurists, such as hues and saturation in an image. This is especially true for neural nets with several hidden layers, also called deep neural networks. One way to think about this is the different ways a computer can understand a digit. If the meaning of a digit is encoded into standard character format, such as UTF-8 or ASCII, the subsequent analysis will be straightforward. However, if the digit is handwritten, such as in the Modified National Standards and Technology database (MNIST), deep layers in a neural network are used to find edges, colors, saturation, and so on necessary for the computer to understand that it is in fact a digit and not something else (LeCunn et al. n.d.). Since ANNs predict based on attributes and not logical reasoning, they sometimes make nonsensical mistakes such as failing to see the difference between dogs and muffins (Yao 2017). While such misclassifications are often humorous, they can sometimes lead to insensitive errors. The autotagging feature of the photo sharing site Flickr ran into a faux pas when it suggested a photo from the Dachau concentration camp be tagged as "jungle gym" (Auerback 2018).

The "synaptic" reasoning inside the network is often so complex that we cannot determine the reason why the network makes a particular connection. Since we don't know the inner connections between the input and output layers, these methods have been called blackbox algorithms. Since futurists have a greater need to understand the reasoning behind the connections and not only just predict individual outcomes, the value of neural networks that use deep learning might be in analyzing a greater variety of data sources. While of limited importance today, as more nontraditional data become available to futurists, deep learning can be useful to process a more diverse pool of data such as image or audio data to look for trends and weak signals.

Are Machines Biased?

Using data analytics to mitigate human bias is not helpful if the data our machines train on have adopted these biases.[4] In fact, without ethical goals and adequate feature engineering, we risk amplifying rather than eliminating the cognitive biases we first set out to combat.

In her book *Weapons of Math Destruction* Cathy O'Neil (2016) points to examples where unprivileged loan applicants have been denied mortgages, where schoolteachers get punished for schools' past performances, and where poor people are microtargeted with predatory schemes culminating from their metadata, such as zip codes and home value. In O'Neill's cases injustices embedded in historic data not only exacerbate but give a veil of false machine-driven objectivity. Social bias has been found in NLP where deep learning algorithms draw word associations when trained on corpora that reflect cultural stereotypes (Bolukbasi et al. 2016). Recently, Amazon had to scrap an AI recruiting tool after it trained on human resources data which demonstrated that women were less likely to get technical jobs in the company (Lavanchy 2018).

However, arguing that algorithmic methods are inherently biased implies that the machines themselves and not the humans feeding them are the sources of this bias, and this is clearly not the case. Several efforts at debiasing are already happening, and computers are being trained to disambiguate unintended word associations. Removing bias from algorithms can be as simple as removing a few input variables or more complex such as neutralizing word embeddings. Algorithms learn meanings and associations by finding the cosine distance between co-occurring terms in word vector spaces, so by deliberately assigning equidistance between stereotypically loaded words and its offensive simile we can greatly reduce these otherwise unfortunate associations. Big tech companies are rolling out initiatives to prevent machine-originated prejudice via programs such as Facebook's Fairness Flow, IBM's AI Fairness 360, and TensorFlow's What-If tools which helps analysts find how a model responds to a single feature, overcoming the problem with blackbox algorithms mentioned earlier (Wiggers 2018, 2019). Google proactively applies sophisticated methods to ensure search results not only yield narrowly relevant queries, but also strive to meet ethical standards (Webster et al. 2018).

But there can also value in studying the outcome of biased algorithmic output for diagnostic reasons. By holding up a "mirror" machine learning models can help organizations learn from deep-rooted, sometimes unconscious biases that affect organizational cultures. Before ditching their AI recruiting system, Amazon hiring managers were able to learn about the hidden biases and language influences that had penalized female candidates in their hiring processes. They learned that successful resumes often contained words like "executed" and "captured" which exuded confidence. These words were more frequently expressed by male candidates (Logg 2019). When training data are cleansed for inherent bias, they are inherently more transparent and reliable in their predictive outcomes than humans.

Machine learning projects that can result in algorithmic bias are in most instances very different from the type of models, data types, and objectives most futurists will encounter in their analytic endeavor. Hence, when futurists control not only the sources of data to be processed but also interpretation of the results, chances of social bias that could skew the analysis are almost nonexistent.

A Digital Wild West, the New Oil Barons, and Privacy Protection

Data-driven foresight can only be as good as the data we have access to. As the saying goes, "garbage in, garbage out." Ironically, while over 2.5-quintillion bytes of data are created every single day, the majority of it is still inaccessible to most people (Ahmad 2018). Moreover, the information discrepancy between those who own and those who don't own data is intensifying. One example is data generated by connected devices. When once dumb gadgets and appliances become smarter, they don't only simplify our lives and help us make economical and efficient choices. Whole supply chains use these data to monitor our domestic habits and behaviors. Such data could be treasure troves for futurists who want to learn about changes in people's lived environment. However, little of these data will likely become accessible to the public anytime soon. One reason is obviously to protect users' privacy. But even when anonymized, companies who use these data to maximize profit for themselves have few incentives to share it.

A leader article in the *Economist* ("The World's Most Valuable Resource Is No Longer Oil, But Data" 2017) urges governments to open up some of their data vaults while recognizing the data economy as public infrastructure. India's digital identity system, Aadhaar, is one such example. Other available data sources are provided by companies that submit datasets to public repositories such as Kaggle to create competitions where companies can draw benefit from the collective knowledge of data scientists. Yet, there is still no widespread available repositories, so the onus is on the individual analyst to not only extract, clean and transform raw data into useable formats, but also understand privacy regulations around the data they want to use for analytical purposes. It is worth mentioning that while futurists can make use of the same tools that allow businesses to micro target individuals, our research is intrinsically disinterested in personally identifiable information. This because our objective is measuring change, not pitching individuals with product offerings. Futurists may therefore find use in anonymized data that are of less value for businesses whose objective is direct customer interaction.

Conclusion

Human intelligence has evolved over millions of years. Machine intelligence for less than a century. Machine intelligence is not superior to that of humans. Data analytic methods are valuable because of how it complements, not replaces human cognition. The same features that make our brain so efficient at deriving meaning are also the qualities that distract us from being thorough and impartial analyst. We can multitask and make sense out of very limited data, but in doing so we make shortcuts and put creative spins which can distort our analysis. We don't always know where our objective analysis ends, and our subjective interpretations start. Data mining and machine learning methods offer the impartiality and analytical capability that humans lack. While full objectivity remains a tenuous goal, data deductive approaches early in the foresight process can help ensure that we base our interpretations on a robust body of unbiased insights. However, while providing more comprehensive data and analytic capabilities, AI is not likely to replace the human capacity to ponder various future outcomes. Instead, it will ensure that we can assess the current situation with a quantitative measurement of precision and accuracy we would not be able to find using human approaches alone.

We can approach the future not with certitude, but with enough granular insights about what happens today to make meaningful and provocative scenarios about the future. We apply data analytics in ways that let us discover big trends and nascent change, and a good understanding of likely cause and effect in closed systems so that we can anticipate what might happen when these systems are affected by autonomous forces in open systems. Data mining helps us do the dirty work of understanding relationships where they exist, but it's up to us to find implications and build narratives on these relationships.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Anne Boysen https://orcid.org/0000-0001-7673-5589

Notes

1. While traditional hypothesis testing uses significance testing to prevent type 1 and type 2 errors, Big Data analytics and machine learning often use different types of validation techniques. This is useful because of the diminishing return of *p* values in large datasets and the ability to test new information against training data.
2. This presumption is based on the absence of such mentions in common foresight discussions. Customer databases may be more frequently used in corporate foresight without being a central topic, which could be due to client agreements and confidentiality constraints.
3. To the extent black swans can be discovered in Big Data, the types of anomaly detections mentioned earlier is likely to be more productive than predictive modeling.
4. It's important to appreciate that the word *bias* usually has a different meaning in machine learning. Bias is a parameter in artificial neural networks which has the same function as the intercept in a linear equation. In this context, bias is used to reference social bias.

References

Ahmad, I. 2018. "How Much Data Is Generated Every Minute? [Infographic]." *Social Media Today*, June 15. https://www.socialmediatoday.com/news/how-much-data-is-generated-every-minute-infographic-1/525692/

Attewell, P. A., and D. Monaghan. 2015. *Data Mining for the Social Sciences: An Introduction*. Oakland: University of California Press.

Auerback. 2018. *Bitwise: A Life in Code*. New York: Pantheon Book.

Bishop, P. 2009. "Horizon Scanning Why Is It So Hard?*"* http://law.uh.edu/faculty/thester/courses/Emerging%20Tech%202011/Horizon%20Scanning.pdf.

Bishop, P., and A. Hines. 2012. *Teaching about the Future*. New York: Palgrave Macmillan.

Bolukbasi, T., K. Chang, J. Zou, V. Saligrama, and A. Kalai. 2016. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings.*" Neural Information Processing Systems Foundation*. https://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf.

Boysen, A. 2019. "How Data Science Enhances Foresight." *Futurist.com by Glen Hiemstra*, February 26. https://www.futurist.com////how-data-science-enhances-foresight-by-anne-boysen/20190226.

"Confirmation Bias and the Power of Disconfirming Evidence." 2017. *Farnam Street*. https://fs.blog/2017/05/confirmation-bias/.

Conway, M. 2006. "An Overview of Foresight Methodologies.*"* http://www.forschungsnetzwerk.at/downloadpub/An-Overview-of-Foresight-Methodologies1.pdf.

Das, T. K., and P. Mohan Kumar. 2013. "Big Data Analytics: A Framework for Unstructured Data Analysis." *International Journal of Engineering and Technology* 5 (1): 153–56.

Dator, J. 2009. "Alternative futures at the Manoa School." *Journal of Futures Studies* 14 (2): 1–18.

Dator, J. 2018. "Emerging Issues Analysis: Because of Graham Molitor." *World Futures Review* 10 (1): 5–10.

Dea, J. 2015. "Do You Know the 4 V's of Big Data?" *INTELEX Blog*, July 16. https://blog.intelex.com/2015/07/16/do-you-know-the-4-vs-of-big-data/.

Gunluk, O., J. Kalagnanam, M. Li, M. Menickelly, and K. Scheinberg. 2019. "Optimal Decision Trees for Categorical Data via Integer Programming." http://www.optimization-online.org/DB_FILE/2018/01/6404.pdf.

Harryson, M., E. Métayer, and H. Sarrazin. 2014. "The strength of 'weak signals.'" *Mckinsey Quarterly*, February . https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-strength-of-weak-signals.

Hiltunen, E. 2008. "Good Sources for Weak Signals: A Global Study of Where Futurists Look For Weak Signals." *Journal of Future Studies* 12 (4): 21–44.

Hiltunen, E. 2010. *Weak Signals in Organizational Futures Learning*. Aalto University School of Economics: Aalto Print. http://epub.lib.aalto.fi/pdf/diss/a365.pdf.

Inayatullah, S. 2009. "Causal Layered Analysis: An Integrative and Transformative Theory and Method." In *Futures Research Methodology, Version 3.0*, edited by J. Glenn, and T. Gordon. Washington, DC: The Millennium Project.

Inayatullah, S. 2014. "Causal Layered Analysis (CLA) Defined (2014)." *Metafuture*. https://www.metafuture.org/causal-layered-analysis-cla-defined-2014/.

Kayser, V., and K. Blind. 2017. “Extending the knowledge base of foresight: The contribution of text mining.” *Technological Forecasting & Social Change* 116:208–15.

Koehrsen, W. 2018. “Beyond Accuracy: Precision and Recall,” towards Data Science.*”* https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c.

Kuntz, M. 2012. “The Postmodern Assault on Science: If All Truths are Equal, who Cares What Science Has to Say?” *EMBO Reports* 13:885–89. doi:10.1038/embor.2012.130.

Lavanchy, M. 2018. “Amazon’s Sexist Hiring Algorithm Could Still Be Better than a Human.” *Phys.org*, November 1. https://phys.org/news/2018-11-amazon-sexist-hiring-algorithm-human.html.

LeCunn, Y., C. Cortes, and C. Burgess. (n.d.). “The MNIST Database of Handwritten Digits.” http://yann.lecun.com/exdb/mnist/.

Linoff, G., and M. J. A. Berry. 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Indianapolis: John Wiley.

Logg, J. M. 2019. “Using Algorithms to Understand the Biases in Your Organization.” *Harvard Business Review*, August 9. https://hbr.org/2019/08/using-algorithms-to-understand-the-biases-in-your-organization.

London School of Economics and Political Science. 2017. “Is Post-Modernism to Blame for our Post-Truth World?” [Audio Podcast], Public Lectures and Events, October 2. http://www.lse.ac.uk/lse-player?id=3892.

Lum, R. 2016. “Trends vs. Emerging Issues: What Is the Difference?” *Visionforesightstrategy.wordpress.com*, April 3. https://visionforesightstrategy.wordpress.com////trends-vs-emerging-issues-what-is-the-difference/20160403.

Molitor, G. T. T. 2017. “The Molitor Model of Change.” *World Futures Review* 10 (1): 13–21. doi:10.1177/1946756717747636.

Mounk, Y. 2018. “What an Audacious Hoax Reveals about Academia.” *The Atlantic*, October 5. https://www.theatlantic.com/ideas/archive/2018/10/new-sokal-hoax/572212/.

Mühlroth, C., and M. Grottke. 2018. “A Systematic Literature Review of Mining Weak Signals and Trends for Corporate Foresight.” *Journal of Business Economics* 88 (5): 643–87.

O’Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pearl, J., and D. Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Popper, K. R. 2002. *The Logic of Scientific Discovery*. New York: Routledge.

Sarniak, R. 2015. “9 Types of Research Bias and How to Avoid Them.” https://www.quirks.com/articles/9-types-of-research-bias-and-how-to-avoid-them

Stephens-Davidowitz, S. 2017. *Everybody Lies: What the Internet Can Tell Us about Who We Really Are*. New York: HarperCollins Publishers.

Taleb, N. N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

“The World’s Most Valuable Resource Is No Longer Oil, But Data.” (2017). *The Economist*, May 6. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

Voros, J. 2003. “A Generic Foresight Process Framework.” *Foresight* 5 (3): 10–21.

Webster, K., V. Axelrod, J. Baldridge, and M. Recasens. 2018. “Mind the GAP: A Balanced Corpus of Gendered Ambiguous Pronouns.” *Transactions of the Association for Computational Linguistics* 6, 605–17. https://arxiv.org/pdf/1810.05201.pdf.

Wiggers, K. 2018. “Google’s What-If Tool for TensorBoard Helps Users Visualize AI Bias.” https://venturebeat.com/2018/09/11/googles-what-if-tool-for-tensorboard-lets-users-visualize-ai-bias/.

Wiggers, K. 2019. “MIT CSAIL Researchers Propose Automated Method for Debiasing AI Algorithms.” https://venturebeat.com/2019/01/26/mit-csail-researchers-propose-automated-method-for-debiasing-ai-algorithms/.

“Word of the Year.” 2016. Oxford Languages. https://languages.oup.com/word-of-the-year/word-of-the-year-2016.

Yao, M. 2017. “Chihuahua or Muffin? Searching for the Best Computer Vision API.” *TOPBOTS*, September 22. https://www.topbots.com/chihuahua-muffin-searching-best-computer-vision-api/.

Author Biography

Anne Boysen holds a Masters Degree in Foresight from the University of Houston and a Graduate Certificate in Business Analytics from Penn State University. She has worked as an analyst in corporate settings, as an independent data analyst and as a foresight consultant in collaboration with other futurists, enriching foresight projects with data analytics and insights. Ms. Boysen teaches data mining for the graduate program in Foresight at the University of Houston.