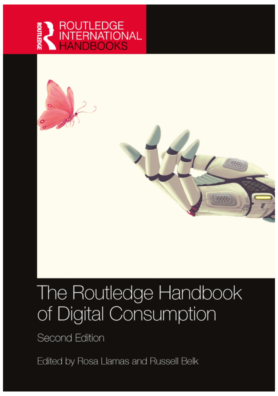

The Routledge Handbook of Digital Consumption

Rosa Llamas, Russell Belk

Critical issues in artificial intelligence algorithms and their implications for digital marketing

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781003317524-16
Elanor Colleoni, Daniela Corsaro
Published online on: 26 Sep 2022

13

CRITICAL ISSUES IN ARTIFICIAL INTELLIGENCE ALGORITHMS AND THEIR IMPLICATIONS FOR DIGITAL MARKETING

Elanor Colleoni and Daniela Corsaro

Introduction

Artificial intelligence algorithms are not just digital technologies for economic developments that require data regulation and governance (Flyverbom *et al.* 2019). They are powerful forces that are reshaping our lives at all levels. These algorithms are used not only in our social life through the recommendation of what to buy, who to connect to, what job to apply to (Ma and Sun 2020), but widely applied also in predicting potential criminals and terrorist attacks (Ganor 2019), judicial sentencing risk scores (Crawford 2013), fraud detection (Ryman-Tubba *et al.* 2018), health assessment (O'Neil 2016), and political campaigning (Nickerson and Todd 2014).

In digital marketing, the application of AI algorithms to the digital traces abundantly disseminated by consumers in contemporary digital environments has improved dramatically the ability of marketers to predict consumer behaviors (De Bruyn *et al.* 2020). As argued by Kotras (2020), this increased predictive ability has revived the old marketers' dream of one-to-one, perfectly adjusted selling techniques at an unprecedented scale. For instance, in everyday life, people rely on recommendations from peers through spoken words, reference letters, news reports, reviews, and so on. Through recommender systems, marketers assist and augment this natural social process to help people sift through information to the point of tailoring customized selling experiences with the goal of increasing their conversion rates. However, as predictive abilities are based on the quantity of the consumer data harvested, extant research has shown how digital marketing has become more and more intrusive and unveiled the mechanisms of a digital environment increasingly based on big-data-based consumer surveillance and the datafication of everyday life (Ball 2017; Cluley and Brown 2015; Zwick and Denegri-Knott 2009). But while marketers have justified this exploitation with the increased relevance of information provided to the consumers and the correspondent reduction of non-relevant data (Darmody and Zwick 2020), several authors have warned about the insidious perils of commercial exploitation of digital information (Beer 2009; Lash and Lury 2007; Thrift 2005; Zwick and Dholakia 2008). Zuboff (2015) has argued that the systematic harvesting, storage, analysis, and commodification of personal data constitute a new form of capitalism that she calls surveillance capitalism. In surveillance capitalism, the

DOI: 10.4324/9781003317524-16

value is created through the real-time public surveillance of the activities and movements of individuals.

More recently, consumer research has focused on the influence marketers exert on consumers when shaping the digital environment that consumers experience in their daily online practices (Ritzer 2015) as a result of AI algorithms and recommendation systems. This loop mechanism (Zuboff 2019) in which algorithms take as input consumers' digital traces (Cluley and Brown 2015) and update the environment accordingly to consumers' predicted needs with "relevant content" (Darmody and Zwick 2020) has been described as a sort of "technological unconscious" (Beer 2009) that, in the long run, is expression of a "production of prediction" (Mackenzie 2015) rather than consumers' choices and tastes (Airoldi 2019). As pointed out by Airoldi (2019), this algorithmic *recursivity* transforms consumers' identities and behaviors.

As observed by Thompson (2019), this increased ability of AI algorithms to predict and shape consumers' reality has produced novel marketing perspectives but also myths regarding how these algorithms have improved marketing efficiency, impact, and returns. In particular, extant research has emphasized the uncontested ability of AI algorithms to portray consumers' desires (Ball 2017; Ritzer 2015) and the resulting power of marketers over consumers (Darmody and Zwick 2020; Zwick and Dholakia 2004, 2008). For instance, Darmody and Zwick (2020) have described marketers as enthusiast about the new data paradigm, and the algorithms and data as allied.

However, extant research in the field of AI ethics is showing that AI algorithms are often biased, and their predictions reflect and reinforce social, economic, and political fractures that can be found in our societies (Crawford 2016; Schroeder 2021). Further digital marketing research is needed to explore the failures of AI algorithms and their impact on consumers and marketers. To date, the inadequacy of AI algorithms to represent consumers properly and in general society has been discussed mainly in relation to the risk of power abuse or fraud associated with it (Jobin *et al.* 2019; Sandvig *et al.* 2014). For instance, authors have shown how artificial intelligence technologies are used to manipulate and persuade citizens, such as through fake news (Gorodnichenko *et al.* 2018), or to manipulate public debate (Bradshaw and Howard 2018). The most prominent example of artificial intelligence manipulating public opinion is probably the one of Cambridge Analytica, a firm that gained access to the personal data of more than 50 million Facebook users and used their psychological profiles in order to target adverts to voters and manipulate their voting intention. But even in these studies, the assumption is that the functional correctness of AI algorithms is distorted by improper behaviors. We lack a critical investigation on the effects of AI algorithms failures resulting from their inner functioning. As pointed out by Crawford (2016), researchers need to explore artificial intelligence algorithms beyond the celebratory thread of the technical discourse that portrays them as powerful objects reflecting and shaping reality. This implies investigating how the algorithms *make sense* of the data and output a *final decision* in a standard context (Pasquale 2015), what is driving its decision, and what are the implications for marketers.

In this chapter, we review the critical issues prompted by algorithmic creation and implications for decision-making process in marketing. The chapter is organized as follows. In the first paragraph, we briefly outline what artificial intelligence is, its usage in marketing and recent developments. In the second paragraph, we map out the critical issues and outline implications for marketing. In the conclusions, we briefly discuss future research needed to address the critical issues associated with artificial intelligence and machine learning development in digital marketing research.

Artificial intelligence and machine learning algorithms: definition, usage, and recent developments

Origins of artificial intelligence

Artificial intelligence represents the broader concept of machines that can think and make decisions autonomously. In the early days of artificial intelligence, the focus was on solving problems that are intellectually difficult for human beings but relatively straight-forward for computers, such as playing chess. These problems were solved by describing a list of formal rules that the computer must follow. However, as it turned out, the real challenge is making the computer performs tasks that are easy for humans but complex in rules, such as recognizing an image (Goodfellow *et al.* 2016). The initial idea has been hard-coding knowledge in formal language using logical inference rules. This approach is known as knowledge base approach. However, none of the projects following this approach led to a major success. This is because human knowledge is largely subjective and intuitive, and therefore very difficult to articulate in a formal way (Goodfellow *et al.* 2016). Artificial intelligence has therefore focused on developing artificial systems that have the ability to acquire their own knowledge. This is the primary goal of machine learning algorithms. Machine learning is an umbrella concept that encompasses all those algorithms devoted to the automated discovery of hidden patterns in large datasets that allow systematic monitoring, segmentation, and prediction of behaviors (Pang and Lee 2008). In machine learning, the algorithm "learns" by defining rules to determine how new inputs will be classified (Mittelstadt *et al.* 2016).

In the last ten years, machine learning applications have grown dramatically, showing great potential in different industries. Prominent example of machine learning algorithm in marketing is recommendation systems. In everyday life, people rely on recommendations from peers through spoken words, reference letters, news reports, reviews, and so on. Recommender systems assist and augment this natural social process to help people sift through information by applying a filter based on affinity between users and items (i.e., homophily). Suggestions for books on Amazon.com or connections on LinkedIn are examples of this customized system. Recommender systems assume that our choices are stable and based on group similarity and imitation; therefore, it is possible to identify similar patterns of behavior among peers based on their relational network (Colleoni, 2012).

AI deep learning revolution

A recent breakthrough in the field of machine learning has been the development of AI deep learning. The idea is to allow computers to learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined through its relation to simpler concepts, creating a deep network with many layers. The hierarchy of concepts enables the computer to learn complicated concepts by building them out of simpler ones (Goodfellow *et al.* 2016). On a conceptual level, deep learning was developed already in 1940, drawing inspiration by the way neurons functions in the brain. McCulloch and Pitts (1943) first came up with the idea that neurons are binary units that activate or deactivate connections based on logical inference, and that the brain is a dynamic system and that neurons in the brain learn by modifying the strength of their connection (LeCun *et al.* 2015). Wiener (1948) formalized these ideas in his famous theory of cybernetics that explains how systems can self-regulate using sensors and feedback loop mechanisms. The idea of neural networks was then abandoned because of the technical problems in implementing it and also because the binary assumption on the neurons' activation was of little help in developing powerful algorithms. In recent years, two key factors have contributed to the resurgence of cybernetics or neural networks algorithms. The first factor is the dramatic increase in computational power. The second and most relevant factor is Hornik's universal approximation theorem (1991) that demonstrates that neural networks containing a sufficient but finite number of neurons can *represent* any continuous function given appropriate weights estimated via recurrent feedback. So that any complex phenomenon we observe in the data can be broken down through a number of simple non-linear functions (i.e., layers), each of them appropriately weighted, whose combination maximizes our ability to predict the phenomenon under observation.

Open questions of AI functioning

To summarize, there are two key points emerging from the big revolution of AI algorithms. First, contrary to standard statistical approaches that apply one single function to the data, in AI algorithms, the data is conceived of as a combination of micro-signals that together can reproduce complex data structure. It follows that the goal of this approach is not to explain the data under observation but to predict it in the best way possible through a number of simple operations. The question comes inevitably of what happens if the data they portray is biased or poorly reflecting the marketplace they aim to replicate. Second, due to its recursive complex nature, while deep learning neural network algorithms can achieve incredible performances, it is basically impossible for humans to understand why and how the algorithms converged to certain parameters and therefore how they made certain decisions or exhibit certain behaviors (Burrell 2016). So, the marketers seem to be losing power over the algorithm choices, rather than being in control (Lorenz *et al.* 2021). In the following, we detail the two key issues outlined above and draw implications for digital marketing.

Artificial intelligence bias and poor representation of reality

AI algorithms are completely reliant on the data they are fed with. In fact, there are simply trying to replicate the patterns observed in the data. It follows that the data used for learning the hidden patterns is crucial in defining what the algorithm will learn and therefore its output decisions. As the algorithm is largely dependent on this data, this process is not neutral and the sample of data used can dramatically change the output.

Bias in knowledge extraction

There are two main techniques used to transform the data into something meaningful: unsupervised and supervised techniques. When using unsupervised techniques, the algorithm autonomously learns the best representation of the data. In this case, the researchers simply feed the algorithm with the data collected, with no filter. When using supervised techniques, the algorithm learns the representation of the data from a pre-trained dataset fed by the data scientists. It has been proven how both these techniques present critical issues that have implications for marketing research and society as a whole (Caliskan *et al.* 2017). In particular, researchers have shown that when working with human behavioral data, such as social media data, if the data contains biases, the algorithm will reproduce these biases (Zhao *et al.* 2017), and it will even amplify them (Douglas 2016). This is true for both techniques in different ways.

Bias in unsupervised techniques

In unsupervised learning techniques, the algorithm learns directly from the data collected with no filter or pre-defined features extraction. In this case, the algorithm learns to mirror the reality it has been exposed to. As unsupervised algorithms learn implicit patterns, several researchers have shown that, when working with human data, the algorithm embeds and, most importantly, reproduces social biases, such as sexism (Basta *et al.* 2019), racism (Chander 2017), gender inequalities (Zhao *et al.* 2018), and, in general, discriminative behaviors toward minorities (Sun *et al.* 2019). For instance, the Microsoft AI chatbot that was left to learn freely surfing the net and ended up generating racist, sexist, and anti-Semitic language in less than 24 hours. Caliskan *et al.* (2017) found that unsupervised algorithms disproportionately associate male terms with science terms and female terms with art terms. Bolukbasi *et al.* (2016) have shown how the associations of words exhibit gender stereotypes, for instance, when looking at occupations, women are associated with sewing and men with carpentry, and again men are associated to computer programmers, while women to homemakers. Douglas (2017) has shown that, when using Google Translate, if one translates "He is a nurse. She is a doctor" to Hungarian and then back again, the genders are switched. These problems apply to all most popular word embedding methods used to analyze big data collection with unsupervised algorithms, such as word2vec (Bolukbasi *et al.* 2016), bert (Bhardwaj *et al.* 2020), and GloVe (Brunet *et al.* 2019). Recently, researchers have shown that these biases appear not only in language-generation, but also in image-generation algorithms. Indeed, Steed and Caliskan (2021) found that feeding the algorithm a cropped photo of a woman, even a famous woman like US Representative Alexandria Ocasio-Cortez and regardless of what she is wearing, the algorithm autocompletes her wearing a low-cut top or bikini 53% of the time.

Bias in supervised techniques

In supervised techniques, the algorithm learns from a pre-trained labeled dataset. In this case, it is the data scientists that provide the input dataset from which the algorithm learns, defining the perimeter of learning for the algorithm. So, for instance, when teaching an algorithm to classify a picture using a supervised technique, the algorithm will be fed with a set of images that are correctly classified according to the final classification. These labeled examples will be used by the algorithm to learn the representation of the data, i.e., the most representative features that distinguish the pictures classified under different categories. As with unsupervised learning, a partial or biased view of the world as represented by the input set can strongly bias the output of the algorithm. This happens when the training sets do not contain an even distribution of subjects in terms of demographic attributes such as race, gender, and age (Xu *et al.* 2021). In 2016, at the top conference in computer science NIPS, Microsoft researcher Crawford presented evidence of how machine-learning-supervised techniques are often biased because they reflect social inequalities and narrow views that are involuntary or voluntary incorporated by the data scientists who create the initial training set on which the algorithm learns. She named the bias "the white-guy problem" following the case of facial recognition technologies which are trained on photos of people who are overwhelmingly white, as they are the engineers who trained the algorithm, and therefore, it has a harder time recognizing non-white faces. As she pointed out, algorithms learn by being fed certain images, often chosen by data scientists, and the system builds a representation of the world based on those images. Several other authors have shown different biases in the training set and how this bias tends to represent and replicate societal fractures and inequalities. Sap *et al.* (2019) presented evidence of how annotators' insensitivity to differences in dialect when creating the training set can lead to racial biases. Koenecke et al. (2020) examined the ability of five state-of-the-art automated speech recognition (ASR) systems, developed by Amazon, Apple, Google, IBM, and Microsoft, to transcribe structured interviews and found that all five ASR systems exhibited substantial racial disparities, with an average word error rate of 0.35 for black speakers compared with 0.19 for white speakers. Using the dermatologist approved Fitzpatrick Skin Type classification system, Buolamwini and Gebru (2018) investigated the racial bias in the cosmetics industry. They evaluated three commercial gender classification systems and found that since the database they are built on overwhelmingly represent of lighter skinned subjects, that darker skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter skinned males is 0.8%. Zhao *et al.* (2017) investigated the impact of gender bias in the food industry and found that the activity of cooking is over 33% more likely to involve females than males in a training set, and a trained model further amplifies the disparity to 68% at test time. West *et al.* (2019) have stressed how the lack of diversity in the high-tech industry is creating a diversity crisis. As a result, the predictive algorithms are simply "replicating patterns of racial and gender bias in ways that can deepen and justify historical inequality" (West *et al.* 2019, p.3).

Bias in feature engineering

Feature engineering refers to the process of extracting relevant features from the data by using domain knowledge. Features are extracted from the data via machine learning techniques or created or imported ad-hoc by the data scientist. In this second case, features are proxies of phenomenon that are assumed to be relevant to reproduce the underlying structure of the data. For instance, when analyzing financial data behavior, one can add an external feature such as the inflation rate during the period under investigation, assuming that the stock market exchanges are related to the inflation rate in a given country. The process of using proxies has been often described as unbiased and completely driven by the algorithm. However, there is growing evidence of the normativity of the choices on the proxy features by those who develop the algorithm (Angwin *et al.* 2016; Bozdag 2013). Indeed, as pointed out by Crawford *et al.* (2014), "datasets are not, and can never be, neutral and theory-free repositories of information waiting to give up their secrets. They require the active interpretation of researchers, all of whom have their own ways of seeing" (Crawford *et al.* 2014, p. 1668). O'Neil (2016), a former Wall Street data scientist, showed how discretional this choice could be and how the use of poor proxies to measure social phenomenon can often be discriminative. For instance, zip code is widely used as one of key variables in credit risk assessment to assess the creditworthiness of individuals. This feature choice ends up in a discriminative bias toward particular ethnicity, who live in clusters in segregated cities. Angwin *et al.* (2016) found that the software used to assess the risk of recidivism in criminals, by using also race as one of the features in the model, was twice as likely to mistakenly flag black defendants as being at a higher risk of committing future crimes and twice incorrect to flag white as low risk. Furthermore, black defendants were 77.3% more likely than white defendants to receive a higher score for violent recidivism. In 2020 in Italy, trade unions in the food delivery industry won a court case against UBER-eats usage of a discriminative algorithm that would reduce the number of deliveries to those workers that were working less hours due to illness or holidays. In sum, features behind algorithms are all but neutral and do have real-world social and political consequences that are just now being questioned.

Implications of AI biases for digital marketing

Data manipulation and interpretation behind algorithms are all but neutral and do have real-world social and political consequences that are just now being questioned. The fact that AI algorithms reproduce biases reflecting societal fractures poses a great danger to society, such as reinforcing damaging stereotypes in downstream applications. However, the negative effects of these biases are not confined to society; they are evident for digital marketing as well and in particular for business activities ranging from product recommendation, targeted advertising, promotions, which strongly rely on these algorithms to segment and recommend products to consumers. Mishra *et al.* (2019) revealed how recommendation systems incorporate cultural biases learned from the reviewers and amplifies them by feeding back through the recommendations. Their study showed that more female images are shown when the search-query is negative (e.g., impulsive shopper) than positive (e.g., sensible shopper). Most importantly, they demonstrated that in the domain of product recommendation, biased recommendations can influence choices made by females. So, to go back to the critical issue raised by Airoldi (2019) about how AI algorithms transform consumer identities, current empirical research indicates that AI-biased predictions can influence consumers toward "some of humanity's worst tendencies" (Floridi and Chiriatti, 2020, p.5).

For digital marketing, AI bias is a critical issue that is increasingly discussed. As pointed out by Desouza *et al.* (2020), there are widespread concerns that these AI failures would generate a "lack of trust" (Desouza *et al.* 2020, p. 6) in AI systems not only among consumers, but most importantly among marketers and in general management, somehow challenging the technological myth of artificial intelligence as the perfect machine (Thompson 2019). To avoid the lack of trust in AI prediction, a number of systems are now being developed (De Bruyn *et al.* 2020). However, to date, we lack a one-fit-all solution.

Artificial intelligence opacity and the role of the marketers

The diffusion of artificial intelligence algorithms based on deep learning has been celebrated as a great success by marketing scholars (Ma and Sun 2020), as these techniques can achieve incredible performances. However, the backdrop of this is that marketers and data scientists have no understanding on why and how the algorithms converged to certain parameters and therefore how they made certain decisions or exhibit certain behaviors. Indeed, deep learning neural networks with multiple layers represent the functional composition of *n* (linear and non-linear) functions. The more the layers, the more complex the system and the less our ability to interpret the results. For instance, the residual network evaluated with a depth net of 152 layers won the first place on the ILSVRC 2015 classification task (He *et al.* 2015). For our human brain, it is already difficult to understand the combined effect of two layers. First wave of machine learning in digital marketing has been mainly driven by interpretable algorithms (Colleoni 2012). For instance, recommender systems used for sponsored content are based on similarity measures whose features can be identified and similarity can be ranked; decision-trees are based on simple decision rules inferred from prior data. Old but widely applied techniques for text mining such as LDA are based on the probability of n-grams occurrences in the text, econometric models, such as regressions, temporal or panel regressions, that are used in marketing attribution models based on utility maximizing decisions. In econometric models, the functional form as well as the variables are also selected according to theory and are included in a way that statistical hypothesis testing can be performed to interpret how they are related (Ma and Sun 2020). In contrast, AI deep learning algorithms result in a black-box, which delivers predictive accuracy, but not interpretive insights, and make parametric statistical hypothesis testing unfeasible, presenting another hurdle for interpretation (Ma and Sun 2020). As pointed out by Burrell (2016), these algorithms are opaque, meaning that "if one is a recipient of the output of the algorithm (the classification decision), rarely does one have any concrete sense of how or why a particular classification has been arrived at from inputs". However, this opacity is not resulting from "corporate secrecy" and willful self-protection by corporations in the name of competitive advantage, as discussed by Pasquale (2015).

Implications of AI opacity for digital marketing

The diffusion of AI algorithms has been celebrated as the highest form of control of marketers over the consumers (Dholakia *et al.* 2020). This control rests on the contradiction on an imagined empowered consumer who consciously shares preferences to the marketer who can control the digital environment to contain the "relevant" set of products for the consumer to choose.

So far, contemporary digital marketers have been portrayed as enthusiast about the new data paradigm, and the algorithms and data as allied (Andrejevic 2014; Beer 2017; Darmody and Zwick 2020). However, the non-interpretability of deep learning algorithms implies that, while marketers have now a set of new powerful algorithms to orient and forge consumer behaviors through hyper-relevance (Darmody and Zwick 2020), they have completely lost control and understanding over the choices made by these algorithms, as there is no transparent model structure and clear linkage between variables. So, more than an increased control of marketers over the consumers, AI algorithms seem to reduce the organizational power of marketers, moving toward an "*algocracy*" (Lorenz *et al.* 2021), where the authority and control are exercised by the algorithmic systems which codify operational, organizational but also professionals rules (Coletta and Kitchin 2017; Danaher *et al.* 2017; Peeters and Schuilenburg 2018).

However, this disconnection between marketing and business and AI decision making is not sustainable for the business. In fact, interpretability is often a deciding factor when applying AI algorithms to products and decision processes (Ma and Sun 2020). Particularly in the field of marketing, the use of reviews to make recommendations to consumers and segmentation purposes relies on the premise that reviews provide a window into consumers' minds that can help the firm better satisfy consumer needs by customizing the offerings. If this cannot be proven, the chances of adoption of an AI algorithm decrease (Ma and Sun 2020). For this reason, a new wave of interpretable AI algorithms to be applied on top of AI algorithms is on their way (Gilpin *et al.* 2018; Ma and Sun 2020). Interpretable AI algorithms provide visibility into how an AI deep learning algorithm makes decisions and predictions and executes its actions (Gilpin *et al.* 2018). The goal of these methods is to explain the rationale for the decision-making process, surface the strengths and weaknesses of the process, and help users to understand the business rationale behind the algorithm predictions.

Conclusion

As stressed by mainstream marketing gurus Kaplan and Haenlein (2019, p.17), AI can be defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation".

The implicit assumption of the "correctness" of AI algorithms and the complete trust in its decisions have characterized marketing adoption of AI so far, in a realization of what Campolo and Crawford (2020) refer to as an "enchanted determinism". However, research mainly in AI ethics and social science is increasingly showing failures of AI systems that have implications not only for society but also for the predictive ability of marketers.

In this chapter, we have shown two AI failures that are impacting the way marketing and marketers make sense of consumers' data. First, AI algorithms are far from being objective and agnostic (Beer 2017), and their unscrutinized application has not only social, economic, and political implications (West *et al.* 2019), but also marketing implications (Mishra *et al.* 2019). Several authors have urged algorithm transparency in the design of the algorithm (Pasquale 2015; West 2019); however, it appears clear that even transparent algorithms can discriminate (Crawford 2016; West *et al.* 2019). The neglected human component is now emerging, showing how the algorithm is profoundly affected by the choices made by the data scientists. As these algorithms are massively used in online platforms for automatically orienting consumers in their journey, the discriminative bias embedded in their output can amplify stereotypes and social fractures (Mishra *et al.* 2019).

Second, the increased opacity over how artificial intelligences learn and adjust their choices implies a loss of control over what algorithms do. The lack of interpretability, along with the impossibility to draw causal conclusions, is perceived as problematic by marketers and it can undermine AI adoption (Ma and Sun 2020; Makariusa *et al.* 2020). It seems that, instead of being marketers allied, AI algorithms require a complete faith in their decisions and no questioning of their choices. While new interpretable algorithms are under development as we speak aiming at providing explanations of AI algorithms, there is a trade-off between AI algorithm interpretability and performances, where complete interpretability means not leveraging on the power of AI deep learning and the full performance means an *algogracy* (Lorenz *et al.* 2021) with little possibility of the marketers and management to understand and control the process.

Further reading

Beer, D. (2017) "The social power of algorithms." *Information, Communication & Society*, 20(1): 1–13.

Crawford, K. (2013) "The hidden biases in big data." *Harvard Business Review.* Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data

O'Neil, C. (2016) *Weapons of Math Destruction*, New York: Crown Books.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M. and Floridi, L. (2021) "The ethics of algorithms: key problems and solutions," *AI & Society,* 37: 1–16.

West, S., Whittaker, M., and Crawford, K. (2019) "Discriminating systems: gender, race and power in AI". *AI Now Institute*. Retrieved from https://ainowinstitute.org/discriminatingsystems.html

References

Airoldi, M. (2019) "Digital traces of taste: methodological pathways for consumer research," *Consumption Markets & Culture*, 24(1): 97–117.

Andrejevic, M. (2014) "The big data divide," *International Journal of Communication,* 8: 1673–1689.

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016) "Machine Bias," *ProPublica 2016*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Ball, K. (2017) "All consuming surveillance: surveillance as marketplace icon," *Consumption Markets & Culture*, 20(2): 95–100.

Basta, B., Costa-jussà, M., and Casas, N. (2019) "Evaluating the underlying gender bias in contextualized word embeddings," *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, P19-3805, 33–39. https://doi.org/10.18653/v1/W19-3805

Beer, D. (2009) "Power through the algorithm? Participatory web cultures and the technological unconscious," *New Media and Society*, 11(6): 985–1002.

Beer, D. (2017) "The social power of algorithms," *Information, Communication & Society*, 20(1): 1–13.

Bhardwaj, R., Majumder, N. and Poria, S. (2020) "Investigating gender bias in BERT," https://arxiv.org/abs/2009.05021.

Bolukbasi, T. Chang, K.W., Zou, J., Saligrama, V., and Kalai, A. (2016) "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings," *Proceedings of the Conference on Neural Information Processing Systems*, 29, 4349–4357.

Bozdag, E. (2013) "Bias in algorithmic filtering and personalization," *Ethics and Information Technology*, 15(3): 209–227.

Bradshaw, S. and Howard, P. (2018) "The global organization of social media disinformation campaigns," *Journal of International Affairs*, 71(1): 23–32.

Brunet, M., Alkalay-Houlihan, C., Anderson, A., and Zemel, R. (2019) "Understanding the origins of bias in word embeddings," *Proceedings of the 36th International Conference on Machine Learning*, 97: 803–811.

Buolamwini, J. and Gebru, T. (2018) "Gender shades: intersectional accuracy disparities in commercial gender classification," *Proceedings of Machine Learning Research,* 81: 1–15.

Burrell, J. (2016) "How the machine 'thinks': understanding opacity in machine learning algorithms," *Big Data & Society,* 3(1): 1–15.

Caliskan, A., Bryson, J. and Narayanan, A. (2017) "Semantics derived automatically from language corpora contain human-like biases," *Science*, 356(6334): 183–186.

Campolo, A. and Crawford, K. (2020) "Enchanted determinism: power without responsibility in artificial intelligence," *Engaging Science, Technology, and Society*, 6: 1–19.

Chander, A. (2017) "The racist algorithm?" *Michigan Law Review,* 115(6): 1023–1045.

Cluley, R. and Brown, S.D. (2015) "The dividualised consumer: sketching the new mask of the consumer," *Journal of Marketing Management*, 31(1–2): 107–122.

Coletta, C., and Kitchin, R. (2017) "Algorhythmic governance. Regulating the 'heartbeat' of a city using the internet of things," *Big Data & Society*, 4(2), 205395171774241.

Colleoni, E. (2012) "New forms of Digital Marketing Research", in Belk, R. and Liama, R. (Eds), *The Routledge Companion to Digital Consumption*, London: Routledge.

Crawford, K. (2013) "The hidden biases in big data," *Harvard Business Review.* Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data.

Crawford, K. (2016) "Can an algorithm be agnostic? Ten scenes from life in calculated publics," *Science, Technology, & Human Values,* 41(1): 77–92.

Crawford, K., Miltner, K., and Gray, M. (2014) "Critiquing big data: politics, ethics, epistemology," *International Journal of Communication,* 8: 1663–1672.

Danaher, J., Hogan, M.J., Noone, C., Kennedy, R., Behan, A., and Paor, A.D. (2017) "Algorithmic governance. Developing a research agenda through the power of collective intelligence," *Big Data & Society*, 4(2), 205395171772655.

Darmody, A. and Zwick, D. (2020) "Manipulate to empower: hyper-relevance and the contradictions of marketing in the age of surveillance capitalism," *Big Data & Society*, 7(1): 1–12. https://doi.org/10.1177/2053951720904112.

De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K.-U., and Von Wangenheim, F. (2020) "Artificial intelligence: pitfalls and opportunities," *Journal of Interactive Marketing*, 51: 91–105.

Desouza, K., Dawson, G., and Chenok, D. (2020) "Designing, developing, and deploying artificial intelligence systems: lessons from and for the public sector," *Business Horizons,* 63(2): 205–213.

Dholakia, N., Darmody, A. Zwick, D., Dholakia, R.R., and First, A.F. (2020) "Consumer choice-making and choicelessness in hyperdigital marketspaces," *Journal of Macromarketing.* 41(1): 65–74, https://doi.org/10.1177/0276146720978257

Douglas, L. (2017) "AI is not just learning our biases; it is amplifying them." Available at https://medium.com/@laurahelendouglas/ai-is-not-just-learning-our-biases-it-is-amplifying-them-4d0dee75931d

Floridi, L. and Chiriatti, M. (2020) "GPT-3: its nature, scope, limits, and consequences," *Minds and Machines,* 30: 681–694.

Flyverbom, M., Deibert, D., and Matten, D. (2019) "The governance of digital technology, big data, and the internet: new roles and responsibilities for business," *Business and Society*, 58(1): 3–19.

Ganor, B. (2019) "Artificial or human: a new era of counterterrorism intelligence?" *Studies in Conflict & Terrorism,* 1: 1–20.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018) "Explaining explanations: an approach to evaluating interpretability of machine learning." Available at https://arxiv.org/abs/1806.00069

Goodfellow, Y., Bengio, Y., and Courville, A. (2016) *Deep learning*. Cambridge: MIT Press.

Gorodnichenko, Y., Pham, T., and Talavera, O. (2018) "Social media, sentiment and public opinions: evidence from# Brexit and# USElection," *National Bureau of Economic Research Technical Report*. http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/ (1 April, 2013).

He, K., Zhang, X., Ren, S., and Sun, J. (2015) "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 770–778.

Hornik, K. (1991) "Approximation capabilities of multilayer feedforward networks," *Neural Networks,* 4(2): 251–257.

Jobin, A., Ienca, M., and Vayena, E. (2019) "The global landscape of AI ethics guidelines," *Nature Machine Intelligence,* 1: 389–399.

Kaplan, A., and Haenlein, M. (2019) "Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons*, 62(1): 15–25.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, JR., Jurafsky, D., and Goel, S. (2020) "Racial Disparities in Automated Speech Recognition." *Proceedings of the National Academy of Sciences*, 117(14): 7684–7689.

Kotras, B. (2020) "Mass personalization: predictive marketing algorithms and the reshaping of consumer knowledge," *Big Data & Society,* 7(2), 1–14.

Lash, S., and Lury, C. (2007) *Global Culture Industry: The Mediation of Things*. Cambridge: Polity.

LeCun, Y., Bengio, Y., and Hinton, G. (2015) "Deep learning," *Nature*, 521(7553): 436–444.

Lorenz, L., Meijer, A., and Schuppan, T. (2021) "The *Algocracy* as a new ideal type for government organizations: predictive policing in Berlin as an empirical case," *Information Polity,* 26(1): 71–86.

Ma, L. and Sun, B. (2020) "Machine learning and AI in marketing – connecting computing power to human insights," *International Journal of Research in Marketing,* 37(3): 481–504.

Mackenzie, A. (2015) "The production of prediction: what does machine learning want?" *European Journal of Cultural Studies*, 18: 429–445.

Makariusa, E., Mukherjeea, D., Foxa, J.D., and Fox, A.K. (2020) "Rising with the machines: a sociotechnical framework for bringing artificial intelligence into the organization," *Journal of Business Research*, 120: 262–273.

McCulloch, W. and Pit, H. (1943) "A logical calculus of ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, 5: 115–130.

Mishra, A., Mishra, H., and Rathee, S. (2019) "Examining the presence of gender bias in customer reviews using word embedding," Working paper. David Eccles School of Business, University of Utah.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016) "The ethics of algorithms: mapping the debate," *Big Data & Society*, 3(2): 1–21.

Nickerson, D. W. and Todd, R. (2014) "Political campaigns and big data," *The Journal of Economic Perspectives*, 28(2): 51–74.

O'Neil, C. (2016) *Weapons of Math Destruction*. New York, NY: Crown Books.

Pang, B., and Lee, L. (2008) *Opinion Mining and Sentiment Analysis*. Boston, MA: Now Publishers.

Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

Peeters, R., and Schuilenburg, M. (2018) "Machine justice. Governing security through the bureaucracy of algorithms," *Information Polity*, 23(3): 267–280.

Ritzer, G. (2015) "Automating prosumption: The decline of the prosumer and the rise of the prosuming machines," *Journal of Consumer Culture,* 15(3): 407–424.

Ryman-Tubba, R., Krauseb, P., and Garnc, W. (2018) "How Artificial Intelligence and machine learning research impacts payment card fraud detection: a survey and industry benchmark," *Engineering Applications of Artificial Intelligence,* 76: 130–157.

Sandvig, C. Hamilton, K., Karahalios, K., and Langbort, C. (2014) "Auditing algorithms: research methods for detecting discrimination on internet platforms," *Proceedings of the International Communication Association*, 1–23.

Sap, M., Card, D., Gabriel, S., Choi, Y., and Smith, N. (2019) "The risk of racial bias in hate speech detection," *Proceedings of the Association for Computational Linguistics (ACL),* P19-1163, 1668–1678.

Schroeder, J. (2021) "Reinscribing gender: social media, algorithms, bias," *Journal of Marketing Management,* 37(3–4): 376–378.

Steed, R. and Caliskan, A. (2021) "Image representations learned with unsupervised pre-training contain human-like biases," *Conference on Fairness, Accountability, and Transparency (FAccT '21),* 3–10.

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J, Mirza, D., Belding, E., Chang, K.W., and Wang, W. (2019) "Mitigating gender bias in natural language processing: literature review," *Proceedings of the Association for Computational Linguistics (ACL),* P19-1163, 1630–1640.

Thompson, C. (2019) "The 'big data' myth and the pitfalls of 'thick data' opportunism: on the need for a different ontology of markets and consumption," *Journal of Marketing Management,* 35(3–4), 207–230.

Thrift, N. J. (2005). *Knowing Capitalism*. London: Sage Publications.

West, S. (2019) "Data capitalism: redefining the logics of surveillance and privacy," *Business & Society*, 58(1), 20–41.

West, S., Whittaker, M., and Crawford, K. (2019) "Discriminating systems: gender, race and power in AI," *AI Now Institute*. Retrieved from https://ainowinstitute.org/discriminatingsystems.html.

Wiener, N. (1948). *Cybernetics*. New York, NY: John Wiley & Sons.

Xu, T., White, J. Kalkan, S., and Gunes, H. (2021) "Investigating bias and fairness in facial expression recognition," *Proceedings of the European Conference on Computer Vision,* 12540, 506–523.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., and Chang, K.W. (2017) "Men also like shopping: reducing gender bias amplification using corpus-level constraints," *Conference on Empirical Methods in Natural Language Processing*, D17-1, 7–11.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., and Chang, K.W. (2018) "Gender bias in coreference resolution: evaluation and debiasing methods," *North American Chapter of the Association for Computational Linguistics (NAACL HTL18)*, 2: 15–20.

Zuboff, S. (2015) "Big other: surveillance capitalism and the prospects of an information civilization," *Journal of Information Technology,* 30(1): 75–89.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

Zwick, D., and Denegri Knott, J. (2009) "Manufacturing customers: the database as new means of production," *Journal of Consumer Culture*, 9(2): 221–247.

Zwick, D., & Dholakia, N. (2004) "Whose identity is it anyway? Consumer representation in the age of database marketing," *Journal of Macromarketing*, 24(1): 31–43.

Zwick, D. and Dholakia, N. (2008) "Infotransformation of markets: introduction to the special issue on marketing and information technology," *Journal of Macromarketing*, 28: 318–325.