Explanations as Discourse: Towards Ethical Big Data Analytics Services

Sadaf Afrashteh
The University of Melbourne
safrashteh@student.unimelb.edu.au

Ida Someh
School of Business,
The University of Queensland

Michael Davern
Faculty of Business and Economics
The University of Melbourne

Abstract

Big data analytics uses algorithms for decision-making and targeting of customers. These algorithms process large-scale data sets and create efficiencies in the decision-making process for organizations but are often incomprehensible to customers and inherently opaque in nature. Recent European Union regulations require that organizations communicate meaningful information to customers on the use of algorithms and the reasons behind decisions made about them. In this paper, we explore the use of explanations in big data analytics services. We rely on discourse ethics to argue that explanations can facilitate a balanced communication between organizations and customers, leading to transparency and trust for customers as well as customer engagement and reduced reputation risks for organizations. We conclude the paper by proposing future empirical research directions.

Keywords: Big data; Analytics services; Ethics; Explanation facilities; Discourse ethics

1 Introduction

Big data analytics has become a prevalent phenomenon in modern society and is a topic of much discussion in the information systems (IS) field. Big data is typically characterized by its large volume, the variety of attributes captured, and the velocity at which the data accumulates, primarily through online interfaces and digital media devices (Constantiou & Kallinikos, 2015). Big data is too diverse and fast-paced to be adequately captured and analysed using traditional data processing tools (Markus & Topi, 2015). In the big data analytics ecosystem, data is harvested through digital devices from even the most mundane activities of an individual's everyday life. The captured data is analysed by algorithms to both tailor products and services to support individuals, and to enable organizational decision making. Big Data Analytics is playing an increasing role in business success, but its effects on both individuals and society are of concern (Newell & Marabelli, 2015; Someh, Davern, Breidbach, & Shanks, 2019).

Algorithmic decision-making that leverages big data analytics to predict customer behaviour is becoming a widespread organizational practice. This decision-making approach is a double-edged sword: with benefits accruing to businesses often at the cost of negative consequences for the individual customers that are the source of the data. As a result, a growing stream of research has begun investigating the ethical and social consequences of big data analytics such as discrimination, de-individualization and lack of privacy (Bélanger & Crossler, 2011). The lack of accountability in the algorithmic decision-making employed in big data analytics means that these negative consequences often continue unchecked (Citron & Pasquale, 2014).

In response to concerns about the consequences of big data analytics, regulators have called for greater transparency in algorithmic decision-making. For example, the recently declared General Data Protection Regulation (GDPR) of the European Union, has introduced a "right to explanation" for the data subject (i.e., the individual whose data is being processed). Specifically, the data subject has the right to get "meaningful information about the logic involved in the process of data analysis" (Goodman & Flaxman, 2016b). However, it is unclear what constitutes "meaningful information" for providing adequate transparency to mitigate the negative ethical and social consequences of big data analytics. In this study, we investigate the role of explanation facilities for algorithmic decision-making as a means for communicating "meaningful information" to data subjects – the users of services enabled by big data analytics. Our research question is thus:

> *How can explanations communicate "meaningful information" to empower and engage users of big data analytics services?*

To address this question, we blend work on discourse ethics and big data analytics with the literature on explanations in intelligent systems. Our goal is to produce a framework to enable us to empirically examine how explanations of big data analytics can effectively communicate "meaningful information" and ultimately mitigate some of the negative ethical consequences of big data analytics.

2 Theoretical Perspectives

2.1 Algorithmic Decision-Making and Ethics

Algorithms can assist decision-making processes through uncovering hidden patterns in data. They can empower decision makers by providing solutions for complex problems (Lepri, Oliver, Letouzé, Pentland, & Vinck, 2018). While algorithmic decision-making provides numerous benefits for individuals, organizations and society, they can also lead to negative and potentially unethical consequences. Benefits of algorithmic decision making include personalised services and products for individuals, innovations and improved decision-making for organizations and improved public health, safety and security at the society level. However, the emerging literature on big data analytics has recognized a variety of unethical consequences of algorithmic decision-making such as discrimination, lack of transparency and lack of accountability. We identify three causes that underlie these ethical issues: the analytic approach, algorithmic model construction, and opaque models.

Algorithmic decision-making employs big data collected from a variety of different sources such as open data sets and customer-generated data. To get the value from the gathered data organizations can employ either an inductive or deductive approach to analysis, although, in big data analytics, the inductive approach is more common. In a deductive approach there is "theory" that drives purposeful data collection and analysis. Ethical issues arise as this "theory" can lead to a confirmation bias in the collection, analysis and interpretation of data (Günther, Rezazade Mehrizi, Huysman, & Feldberg, 2017).

In an inductive approach, data is often collected without a specific intended purpose and analysed to generate "insights" – patterns in the data that can guide action. The patterns however may simply be statistical artefacts like spurious correlations from which causal predictions are often erroneously imputed. Clustering algorithms are often applied in this inductive manner to classify, profile and categorize individuals, either directly or indirectly, based on sensitive attributes like gender, religion and age, resulting in discrimination. Given the inductive approach, individuals may be classified as belonging to a group they themselves do not identify with – in what is known as de-individualization. The potential ethical consequences that can arise from both inductive and deductive analytic approaches, is exacerbated by the lack of power and influence the data subject has over of the analysis, relative to the data controller responsible for directing the analysis.

Algorithmic decision-making works by building a model from historical training data that allows it to then be applied to future decisions, whether through predictions, classifications or other action. This machine learning approach to model construction can be ethically problematic in three ways: (1) incorrect labels (2) sampling bias and (3) incomplete data. First, through differences between labels applied to attributes in the training data and the data set to which the model is to be applied. To the extent that the attribute labels and definitions from the training data are incorrect or different to the context of application, the validity of the model is brought into question. As a result, individuals may be misclassified or otherwise misdealt with. In a similar vein, a biased sample data set can lead to a model that is discriminatory. For example, consider the modelling of suburban crime rates. If the historical training data comprises a disproportionate number of suburbs with a high crime rate, the model may be biased and suggest inappropriate actions or interpretation in settings not adequately represented in the training data (i.e., lower crime rate suburbs). Finally, incomplete training, where important characteristics of the actual population are at best only partially present can lead to a model of questionable validity. Somewhat ironically, this can occur out of an ethical concern to preserve individual data privacy, but it may cause ethically questionable algorithmic decision-making.

Despite the benefits of algorithmic decision-making, most algorithms function as black-boxes. In the context of big data analytics, this is often because the way the machine learning algorithm learns is not understandable by humans and so it is inherently an opaque decision process (Sinha & Swearingen, 2002). Even when there is a human interpretable model and algorithm, organizations can as a matter of policy not disclose the workings of the algorithm to internal users, let alone the subjects of the data (e.g., for competitive reasons). Whether inherent or by policy, the opacity of algorithmic decision-making can lead users and data subjects to distrust the algorithmic model and its decision (Ribeiro, Singh, & Guestrin, 2016). More importantly, to the extent that algorithmic decision-making is opaque, it is hard to ensure full and proper accountability for the decisions made. This has led to calls for improving transparency of algorithmic decision-making. Transparency enables greater accountability for decisions made, thereby enhancing trust (Lepri et al., 2018), but is particularly problematic when the machine learning methods employed are opaque by nature.

2.2 Discourse Ethics

Ethical issues have long been a concern in IS (Markus & Topi, 2015; Martin, 2015) and a variety of theoretical approaches have been used to explore ethics in the IS context. We follow Mingers and Walsham (2010) and use Habermas' discourse ethics as our theoretical approach. We adopt a discourse ethics approach because, unlike other ethics theories, discourse ethics focuses on the communication between the stakeholders as the process by which ethical outcomes are achieved (as opposed to focusing on defining ethical outcomes per se). This focus on the process of communicative action lends itself to investigating our research question on the communication of "meaningful information" to users. Central to discourse ethics is the notion of an ideal speech situation, in which stakeholders engage as equals in the discourse, there is no coercion, and stakeholders have opportunity to question the claims of others, and to present their own claims and needs (Mingers & Walsham, 2010).

Within the algorithmic decision-making context, ethical discourse will emerge when users and decision-makers experience the ideal speech situation. To achieve this aim, users' need to understand the underlying logic of the decision-making process. Introducing transparency opens a lens to reach a fairer communication among stakeholders by providing users with greater clarity about the application of algorithmic decision-making models in the services they use (Lepri et al., 2018). System generated explanations can be used as a means to enhance transparency. There is a well-established literature in IS on how such explanations can shed light on the logic and conclusions made by decision-making technologies and enhance ease of use and trust (Gregor & Benbasat, 1999). When users of big data analytics services are better informed about the processes of algorithmic decision making with big data, they are more likely to engage with decision-makers (Günther et al., 2017) and the resulting discourse may move closer to the ideal speech situation.

2.3 Explanations in intelligent systems

The ability of systems to provide explanations of system decisions or recommendations has long been seen as critical to the acceptance of the decisions and recommendations by users (Dhaliwal & Benbasat, 1996; Hayes-Roth & Jacobstein, 1994; Ye & Johnson, 1995). From the outset explanations have been defined as providing useful descriptions of why requested data was needed, and the reasoning employed in processing that data, and the basis for any recommendation or decision (Clancey, 1983). Explanations have been found to help users in three different ways: (1) to cater for user requirements when their expectations are not fulfilled, (2) to facilitate learning and (3) to deliver information required for problem-solving and decision-making (Gregor & Benbasat, 1999).

We conducted a systematic review of the IS literature focused on explanations using the following search criteria. First, we chose as a foundation, Gregor and Benbasat's (1999), synthesis of the explanation literature, and searched forward for papers that cite this foundation work. We chose this paper as our foundation, because it provided a highly cited, comprehensive synthesis of the explanation literature, and it historically coincided with a time when data driven approaches (data mining, knowledge discovery, and what would now be called data analytics) were just beginning to garner traction in IS practice (in large part due to the increased availability of machine readable data brought about by the boom in e-commerce). Second, we limited our search to papers published in AIS Basket of Eight Journals. Third, we selected papers that used "Explanation" or "Explainability" in their abstract. Fourth, we chose papers with an empirical component that were published from 1999 to 2018. Five, papers before 2007 with less than 25 citation counts were excluded. Our search resulted in 10 papers relevant to our study, as shown in Table 1.

Explanations are used across a range of intelligent systems such as Knowledge-based Systems (KBS) (Arnold, Clark, Collier, Leech, & Sutton, 2006; Dhaliwal & Benbasat, 1996; Gregor, 2001; Gregor & Benbasat, 1999; Mao & Benbasat, 2000; Smedley & Sutton, 2007), Decision Support Systems (DSS) (Berendt & Preibusch, 2014; Gönül, Önkal, & Lawrence, 2006; Tan, Tan, & Teo, 2012), Expert Systems (Arnold et al., 2006; Bohanec, Kljajić Borštnar, & Robnik-Šikonja, 2017; Ye & Johnson, 1995), Recommendation Agents (RA) (McSherry, 2005; Sinha & Swearingen, 2002; Wang & Benbasat, 2007) and case-based reasoning systems (Sørmo, Cassens, & Aamodt, 2005). In all these cases, both machines and individuals play a part in the decision-making and problem-solving process. Explanations assist human decision makers by reducing their cognitive load in the decision-making process as well as enhancing the quality of decisions and recommendations made (Gregor, 2001). From the users' perspective, explanations positively influence perceptions of the system and can also help users learn in the problem domain (Gregor & Benbasat, 1999), primarily by enhancing transparency. In the consumer facing context of recommendation agents, this transparency helps build customer trust in system recommendations by exposing the logic of the underlying the processes to the customer (Sinha & Swearingen, 2002).

A key issue in the prior literature is the role of expertise of the user in the application domain. The effects of explanations vary significantly depending on the expertise of the user. Feedforward explanations conveying declarative knowledge are more useful for novices as they provide information on the system's inputs and their relationships. However, experts prefer feedback explanations that facilitate transfer of procedural knowledge. Feedback explanations provide information on how a decision has been made (Arnold et al., 2006). In the context of big data analytics services, the expertise of the user – the data subject – may be unknown but is likely skewed toward the novice end of the spectrum.

The content of explanations has been categorized into a taxonomy of four different types of information provided to users: *Terminology, Trace, Justification and Strategy* (Gregor & Benbasat, 1999). Terminological explanations are the "knowledge of the concepts and relationships of a domain that experts use to communicate with one another" (Swartout & Smoliar, 1987). Trace explanations describe, ex post, the actual reasoning process employed by the system to reach a specific decision or conclusion. Justification explanations provide the basis, often in terms of domain specific knowledge, for the conclusion or recommend made by a system. Strategy explanations explain the goals of the system that underlie its reasoning process (Sørmo et al., 2005).

A parallel taxonomy to that of Gregor and Benbasat's (1999) comes from the literature on decisional guidance, which is defined as how a system enlightens or persuades users in their decision making (Silver 1991). Conceptually Silver distinguishes between informative guidance which seeks to enlighten users, and suggestive guidance which seeks to sway users (Silver, 2006). Informative guidance is somewhat akin to terminological explanations but there is not a clear mapping between the two taxonomies and decisional guidance has been operationalised in a variety of ways (Davern & Parkes 2010). Pragmatically, we chose to remain with the Gregor and Benbasat (1999) taxonomy as it is arguably more directly translatable into features in an artefact, consistent with the design science approach.

An important aspect of providing explanations is how to communicate them to users. Prior literature reveals a variety of different mechanisms have been employed to provide explanations, including automated provision, user-invoked and intelligent. Automatic explanations are always available without the users' request. Conversely, user-invoked explanations are presented only at the request of the user. Alternatively, the provision of explanations itself can employ intelligent approaches to tailor the response to a particular user or type of user based on some underlying model of the user. At a more pragmatic level there are also choices in the presentation format of the explanations, for example textual, audio-visual, animation and other multimedia techniques. At issue here in both the presentation format and provision mechanisms is interactivity, one the key cognitive qualities of information systems (Davern et al 2012).

The guidance provided by explanations serves to inform and persuade users (Silver, 2006). Toulmin's model of argumentation has been the primary approach in the literature to exploring the persuasiveness of explanations (Ye & Johnson, 1995). This model identifies six elements of argumentation: *Claims, Data, Warrants, Backing, Qualifiers* and possible *Rebuttals*. A claim is a state which is proposed to be accepted. Data is the basis for the argument. Warrants provide the connection from data to the claim, thereby justifying the claim. Backing is concerned with supporting the trustworthiness of warrants in case their validity is doubted. To incorporate issues of the degree of certainty of a claim, qualifiers are used. Possible rebuttals indicate conditions in which the warrant is not applicable and as a result the conclusion can be overturned. Toulmin's model provides a useful structure of the different components to an explanation (i.e., is the explanation data, backing, qualifier etc.) (Gregor & Benbasat, 1999).

A broader literature on explanations has also emerged in the computer science discipline (Letham, Rudin, McCormick, & Madigan, 2015; Ustun & Rudin, 2016), frequently referred to as Explainable Artificial Intelligence (XAI). AI systems autonomously learn from data and mimic human behaviour. The XAI literature focuses on the design of AI models that are inherently explainable while maintaining high learning performance of the underlying algorithms. One example is the Defence Advanced Research Project Agency (DARPA)'s XAI initiative that aims to develop new AI systems that can explain their rationale, determine their strengths and weaknesses, and provide an understanding of how they will act in the future. Researchers have also developed tools such as LIME (Local Interpretable Model-Agnostic Explanations) that produce instance-specific explanations for the outputs of any classifier algorithms (Ribeiro, Singh, & Guestrin, 2016; Valenzuela-Escárcega, Nagesh, & Surdeanu, 2018). Tools like LIME focus on trying to ex post explain, an otherwise opaque, analytics result, rather than the actual reasoning processes employed in getting the result. While this can be useful, it is not aimed at providing meaningful information that could be communicated with the customers of big data analytics services.

Study	Theoretical foundation	Method/Context	Independent Variables	Dependent Variable	Findings
(Arnold et al., 2006)	Toulmin's model of argument, Adoptive Control of Thought-Rational theory	Experiment in insolvency	Existence of explanation, Types of explanation	- Adherence - Explanation accesses	Novices use feedforward explanations to acquire declarative knowledge while experts use feedback explanations to acquire procedural knowledge. Feedback enhances adherence to the recommendation.
(Gregor, 2001)	- Cognitive effort - Cognitive learning	Experiment in cooperative problem solving	Systems with and without requirement of cooperative problem solving	- Use of explanation - Problem solving performance	- A requirement for cooperative problem solving was associated with greater use of explanations. - Positive relationship between explanations and improved performance was more noticeable when problems requiring cooperation were undertaken. - The frequency of use of explanations in total was positively related to problem-solving performance.
(Limayem & DeSanctis, 2000)	Theory of breakpoints in group interaction	Experiment in group decision support system context	Objective variables, Perceptions of the group decision process and outcomes, Perceptions of the multicriteria decision making models and group decision system support	- Model understanding	- System explanations can improve decisional outcomes due to improvement in user understanding of decision models
(Wang & Benbasat, 2007)	Theory of interpersonal communication	Experiment in e-commerce	Types of explanation as how, why and trade-off explanations	- Three beliefs in trust as benevolence, competence and Integrity	- The use of how explanations increases users' competence and benevolence beliefs. - The use of why explanations increases their benevolence beliefs. - The use of trade-off explanations increases their integrity beliefs.
(Mao & Benbasat, 2000)	Cognitive effort theory, Novice-expert differences theories, Question asking theories	Experiment in financial analysis	Level of users' expertise	- Nature of explanation use - Explanation types	- Experienced professionals requested more trace and fewer justification and strategic explanations than novices. - Experts' knowledge makes it easier for them to detect anomalies in KBS output.

Study	Theoretical foundation	Method/Context	Independent Variables	Dependent Variable	Findings
(Tan et al., 2012)	Cognitive effort theory	Experiment in customer decision aid context	Explanation-featured decision aid (Trace, Justification, Strategy)	- perceived decision confidence - decision time - decision quality	More elaborated explanation aid could increase a consumer's decision confidence leading to less cognitive effort and lower product choice made.
(Li & Gregor, 2011)	Theory of explanations	Experiment in online advisory services	Service type (with or without explanation)	- Level of users' decision process satisfaction - Level of decision-advisory transparency	- Different types of self-assessment tools lead to different levels of decision support effectiveness, measured in terms of decision process satisfaction and decision-advice transparency. - Decision-advice transparency is shown to have a stronger influence over empowerment outcomes.
(Giboney, Brown, Lowry, & Nunamaker Jr, 2015)	- Cognitive fit theory - Person-environment fit paradigm	Experiment in deception detection	- Explanation quality - Perceived usefulness of KBS -Perceived ease of use - Explanation cognitive fit	- Perceived usefulness of KBS - Explanation quality - Explanation evaluation time - Explanation influence	- Explanation quality will increase explanation influence and perceived usefulness of the KBS. - Perceived usefulness of the KBS will increase explanation influence. - Perceived ease of use will increase perceived usefulness. - Explanation cognitive fit will increase explanation quality and explanation evaluation time.
(Martens & Provost, 2014)	- Three-gap framework for explanations	Case study in web page classification	- Global explanations Instance-level explanations	- Explanation performance	- Global explanations in the form of a decision tree or a list of the most indicative words do not provide a satisfactory solution.
(Rader, Cotter, & Cho, 2018)		Experiment in social media	Types of explanations	- Awareness - Correctness - Interpretability - Accountability	- The What, How, and Why explanations all supported both awareness and accountability. - Only the what explanations supported correctness. - Only the how explanations supported interpretability.

Table 1. Explanations in Information Systems Literature

3 A Framework for Explanations in Big Data Analytics

We develop a research framework, presented in Figure 1, that illustrates how explanation facilities can be used to communicate meaningful information about algorithmic decision making to external customers. The communication occurs through a discourse process (Habermas, 1984) and can result in both customer-related and organizational outcomes. Our framework consists of four components: Explanation Facility, Ethical Discourse, Customer Outcomes, and Organizational Outcomes. The framework and the definition of concepts follows.

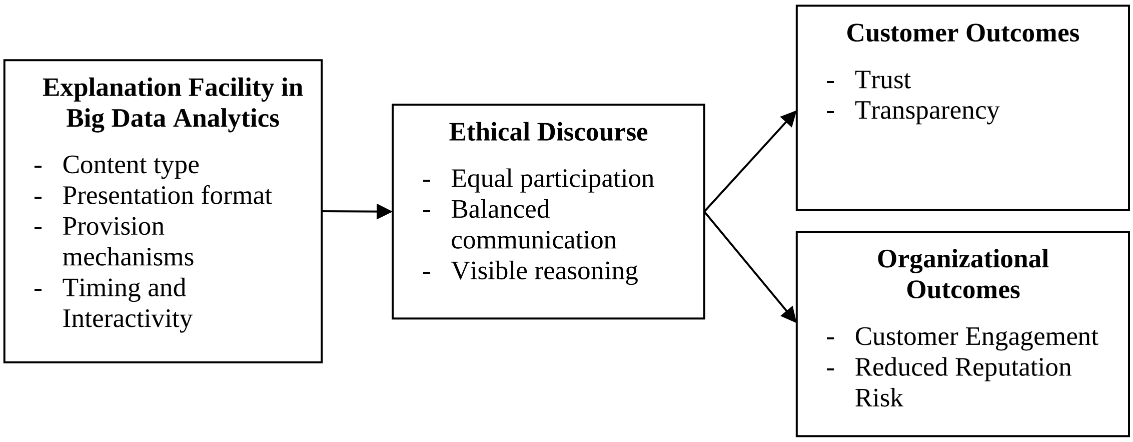

Figure 1: A Framework for Explanations in Big Data Analytics

3.1 The Explanation Facility in Big Data Analytics

The explanation facility seeks to provide meaningful information to external customers about analytics and decisions made within a big data service. Extrapolating from the prior work on explanations, we define three components of explanation facilities in big data analytics: content type, interactivity and timing. Content type follows Gregor and Benbasat's (1999) explanations taxonomy of terminology, trace, justification and strategic. Interactivity captures aspects of presentation format and provision mechanism from the prior literature but recognizes a broader range of means of engaging with users to provide explanations (e.g., interactive and immersive visualizations) (Davern et al 2012). Timing of explanations in big data analytics draws on the feedback and feedforward distinction (Arnold et al., 2006; Ye & Johnson, 1995) and recognizes explanations can vary in what and when they are provided. For example ex ante what data is required and why, or ex post what data was used and how it was analysed.

By providing meaningful information to customers of big data analytics services, the explanation facility can raise awareness and help empower customers in their interaction with service providers.

3.2 Ethical Discourse

Discourse process refers to the communication and engagement between service providers and customers that is enabled and influenced by the provision of explanations. Since the goal is to nudge the discourse towards the ideal speech situation (Habermas, 1984; Mingers &

Walsham, 2010) we conceptualize the discourse as having three components: equal participation, balanced communication, and visible reasoning. Equal participation requires a level playing field in the interaction where the service provider does not exploit a power differential to coerce the customer into providing data or making particular choices. Balanced communication means that customers (and indeed organizations) can freely question and negotiate to achieve an agreement on decision-making processes and data usage. A balanced communication and negotiation between customers and organizations can create a win-win situation for both stakeholders (Günther et al., 2017). As part of ensuring balanced communication and levelling the playing field, the reasoning processes of the algorithmic decision-making needs to be made visible. Conceptually, this could be done by explicating the different elements of Toulmin's argumentation model as part of the discourse. We propose that an appropriately designed explanation facility (i.e., comprising the three components we have identified) will enable an ethical discourse between service providers and customers.

Proposition 1: *An appropriately designed explanation facility in big data analytics services will enable an ethical discourse between service providers and customers.*

3.3 Customer Outcomes

Customer outcomes are the benefits to customers of providing explanations and enabling an appropriate discourse. Transparency and trust are two key outcomes for customers. Transparency emerges when decision-making processes and outcomes are explained to users in a way that they can understand (de Laat, 2018; Mittelstadt, 2016; Sinha & Swearingen, 2002). Explanation facilities and ethical discourse can also help individuals build more trust in a service when they know there is transparency and they can understand why certain decisions are made (Lepri et al., 2018; Ribeiro et al., 2016; Sinha & Swearingen, 2002; Wang & Benbasat, 2007). As such, we propose that:

Proposition 2*: An ethical discourse between service providers and customers will lead to outcomes beneficial to customers such as trust and transparency.*

3.4 Organizational Outcomes

For organizations, explanations to customers will positively influence customers' perceptions about the organization. The shared understanding between organizations and customers about the logic underlying the decision-making process will lead to better customer engagement (Günther et al., 2017). More broadly, by enabling an appropriate discourse process, explanations can help ensure customers and organizations have consistent expectations, and the actions of organizations are consistent with those expectations. This substantially reduces the organization's reputation risk in the use of big data analytics (Silver, 2006). Thus, we propose that:

Proposition 3: *An ethical discourse between service providers and customers will lead to beneficial outcomes to organizations such as customer engagement and reduced reputation risk.*

4 Conclusion and Future Research

In this paper, we reviewed literature on explanations and utilized discourse ethics to propose a framework with a set of propositions for the use of explanations in big data analytics services. Our framework provides insight into the use of explanations to address ethical issues with big data analytics, and more pragmatically contributes to understanding what could constitute "meaningful information" for users as required by the GDPR (Goodman & Flaxman, 2016a).

The framework shows that appropriately designed explanation facilities gives rise to an ethical discourse in which organizations and customers equally participate and negotiate decision-making processes and outcomes that affect individuals. By providing clarity and transparency to users about the process and conclusions of algorithmic decision-making such a discourse can enhance customer trust, and engagement (Bussone, Stumpf, & O'Sullivan, 2015; Gregor & Benbasat, 1999; Lepri et al., 2018).

While prior research in intelligent systems and explainable AI have explored the design of explanation facilities the focus has been primarily on improving decision-making or reliance on systems by organizational decision makers (Gregor, 2001; Tan et al., 2012; Wang & Benbasat, 2007). The context of big data analytics enabled services and the regulatory requirement for "meaningful information" to be provided to data subjects brings to the fore the ethical dimension. We built on the prior literature to address this concern, and in integrating in the theoretical lens of discourse ethics consider explanations not just as a means for improving decision quality, but also for enabling ethical discourse – and shifting that discourse towards the notion of ideal speech. While we consider the various design, elements identified in prior research, in expanding the goal beyond decision quality to include enabling an ethical discourse we expect will lead to quite different developments of even established design elements.

More broadly our approach emphasizes the important role of explanations as communication tools, which brings into the mix new theories in ethical discourse and opens the possibility for drawing theoretical insight from the established literature on organizational communication enabled by IT artefacts (Te'eni, 2001). We look to future research to elaborate on and empirically test our propositions and explore the role of explanations in enabling ethical discourse to the consequent benefit of a broader range of stakeholders (i.e.,. both organizations and customers).

From the practical perspective, our proposed framework provides a guide for organizations on how they can design their explanation facilities to ethically benefit from deploying big data analytics services. Moreover, given the discourse ethics lens, they can be guided how to develop their communication strategies in their interactions with their customers. We look forward with anticipation to seeing how the design of explanation facilities evolves in practice as organization increasingly adopt, either voluntarily or through regulation, an ethical lens on design.

Further research is required to understand how organizations can formulate discourse strategies and tools. An example of these strategies can be designing social bots as explanations interface. These bots can provide transparency with presenting explanations about systems' outputs to customers through which customers can be informed about why the results are reached and interactively share how they feel afterwards. Also, customers can provide organizations with their feedback about systems' performance and this will result in improving the systems as well. In this regard, further studies can be done exploring which types of explanations are more effective leading to greater transparency and customers' engagement in discourse to reach the level playing field situation.

References

Arnold, V., Clark, N., Collier, P. A., Leech, S. A., & Sutton, S. G. (2006). The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions. *MIS Quarterly*, 30(1), 79-97. doi:10.2307/25148718.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: a review of information privacy research in information systems. *MIS Quarterly, 35*(4), 1017-1042. doi:10.2307/41409971.

Berendt, B., & Preibusch, S. (2014). Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence. *Artificial Intelligence and Law, 22*(2), 175-209. doi:10.1007/s10506-013-9152-0.

Bohanec, M., Kljajić Borštnar, M., & Robnik-Šikonja, M. (2017). Explaining machine learning models in sales predictions. *Expert Systems with Applications, 71*, 416-428. doi:10.1016/j.eswa.2016.11.010.

Bussone, A., Stumpf, S., & O'Sullivan, D. (2015). *The role of explanations on trust and reliance in clinical decision support systems.* Paper presented at the Healthcare Informatics (ICHI), 2015 International Conference.

Citron, D. K., & Pasquale, F. (2014). The scored society: due process for automated predictions. *Washington Law Review, 89*, 1.

Clancey, W.J. (1983). The epistemology of a rule-based expert system – a framework for explanation. *Artificial Intelligence*, 20(3), 215-251.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology, 30*(1), 44-57. doi:10.1057/jit.2014.17

Davern, M.J. & Parkes, A. (2010). Incommensurability in design science: which comes first – theory or artefact? In D. Hart & S. Gregor (Eds.), *Information Systems Foundations: The role of Design Science*, 75-90: ANU e-press.

Davern, M.J., Shaft, T. & Te'eni, D. (2012) Cognition Matters: Enduring Questions in Cognitive IS Research. *Journal of the Association for Information Systems*, 13(4), article 1.

de Laat, P. B. (2018). Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability? *Philosophy & Technology* 31(4), 525-541. doi:10.1007/s13347-017-0293-z

Dhaliwal, J. S., & Benbasat, I. (1996). The use and effects of knowledge-based system explanations: theoretical foundations and a framework for empirical evaluation. *Information Systems Research, 7*(3), 342-362. doi:10.1287/isre.7.3.342

Giboney, J. S., Brown, S. A., Lowry, P. B., & Nunamaker Jr, J. F. (2015). User acceptance of knowledge-based system recommendations: Explanations, arguments, and fit. *Decision Support Systems, 72*, 1-10.

Gönül, M. S., Önkal, D., & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems, 42*(3), 1481-1493. doi:10.1016/j.dss.2005.12.003

Goodman, B., & Flaxman, S. (2016a). *EU regulations on algorithmic decision-making and a "right to explanation"*. Paper presented at the ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. http://arxiv. org/abs/1606.08813 v1.

Goodman, B., & Flaxman, S. (2016b). European Union regulations on algorithmic decision-making and a" right to explanation". *arXiv preprint arXiv:1606.08813*. doi:10.1609/aimag.v38i3.2741

Gregor, S. (2001). Explanations from knowledge-based systems and cooperative problem solving: an empirical study. *International Journal of Human-Computer Studies, 54*(1), 81-105. doi:10.1006/ijhc.2000.0432

Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 497-530. doi:10.2307/249487

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems, 26*(3), 191-209. doi:10.1016/j.jsis.2017.07.003

Habermas, J. (1984). *The theory of communicative action* (Vol. 2): Beacon press.

Hayes-Roth, F., & Jacobstein, N. (1994). The state of knowledge-based systems. *Communications of the ACM, 37*(3), 26-39.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes. *Philosophy & Technology*. 31(4), 611-627. doi:10.1007/s13347-017-0279-x

Li, M., & Gregor, S. (2011). Outcomes of effective explanations: Empowering citizens through online advice. *Decision Support Systems, 52*(1), 119-132. doi:10.1016/j.dss.2011.06.001

Limayem, M., & DeSanctis, G. (2000). Providing Decisional Guidance for Multicriteria Decision Making in Groups. *Information Systems Research, 11*(4), 386-401. doi:10.1287/isre.11.4.386.11874

Mao, J., & Benbasat, I. (2000). The use of explanations in knowledge-based systems: Cognitive perspectives and a process-tracing analysis. *Journal of Management Information Systems*, 17(2), 153-179. doi:10.1080/07421222.2000.11045646

Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business.* Paper presented at the Report on a Research Agenda Setting Workshop. Bentley University.

Martens, D., & Provost, F. (2014). Explaining data-driven document classifications. 38(1): 73-99. doi:10.25300/misq/2014/38.1.04

Martin, K. E. (2015). Ethical issues in the big data industry. *MIS Quarterly Executive, 14(2),*, 67-85.

McSherry, D. (2005). Explanation in Recommender Systems. *Artificial Intelligence Review, 24*(2), 179-197. doi:10.1007/s10462-005-4612-x

Mingers, J., & Walsham, G. (2010). Toward ethical information systems: the contribution of discourse ethics. *MIS Quarterly, 34*(4), 833-854. doi:10.2307/25750707

Mittelstadt, B. (2016). Automation, Algorithms, and Politics| Auditing for Transparency in Content Personalization Systems. *International Journal of Communication, 10*, 12.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems, 24*(1), 3-14. doi:10.2139/ssrn.2644093

Rader, E., Cotter, K., & Cho, J. (2018). *Explanations as mechanisms for supporting algorithmic transparency.* Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?". *Proceedings of KDD'16:* 1135-1144. doi:10.1145/2939672.2939778

Silver, M. (1991). Decisional guidance for computer-based decision support. *MIS Quarterly*, 15(1), 105-122.

Silver, M. (2006). Decisional Guidance: Broadening the Scope *Human Computer Interaction and Management Information Systems: Foundations*: M.E. Sharpe.

Sinha, R., & Swearingen, K. (2002). *The role of transparency in recommender systems.* Paper presented at the CHI'02 extended abstracts on Human factors in computing systems.

Smedley, G., & Sutton, S. G. (2007). The effect of alternative procedural explanation types on procedural knowledge acquisition during knowledge-based systems use. *Journal of Information Systems, 21*(1), 27-51. doi:10.2308/jis.2007.21.1.27

Someh, I., Davern, M., Breidbach, C. F., & Shanks, G. (2019). Ethical issues in big data analytics: A stakeholder perspective. *Communications of the Association for Information Systems, 44*(1), 34.

Sørmo, F., Cassens, J., & Aamodt, A. (2005). Explanation in Case-Based Reasoning–Perspectives and Goals. *Artificial Intelligence Review, 24*(2), 109-143. doi:10.1007/s10462-005-4607-7

Swartout, W. R., & Smoliar, S. W. (1987). On making expert systems more like experts. *Expert Systems, 4*(3), 196-208. doi:10.1111/j.1468-0394.1987.tb00143.x

Tan, W.-K., Tan, C.-H., & Teo, H.-H. (2012). Consumer-based decision aid that explains which to buy: Decision confirmation or overconfidence bias? *Decision Support Systems, 53*(1), 127-141. doi:10.1016/j.dss.2011.12.010

Te'eni, D. (2001). A cognitive affective model of organizational communication for designing IT. *MIS Quarterly*, 25(2), 251-312.

Ustun, B. & Rudin, C. (2016). Supersparse linear integer models for optimized medical scoring systems. *Machine Learning*, 102(3), 349-391.

Wang, W., & Benbasat, I. (2007). Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs. *Journal of Management Information Systems, 23*(4), 217-246. doi:10.2753/mis0742-1222230410

Ye, L. R., & Johnson, P. E. (1995). The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly*, 157-172. doi:10.2307/249686

Zupon, A., Alexeeva, M., Valenzuela-Escárcega, M., Nagesh, A., & Surdeanu, M. (2019, June). Lightly-supervised Representation Learning with Global Interpretability. In *Proceedings of the Third Workshop on Structured Prediction for NLP*, 18-28.

doi: https://doi.org/10.3127/ajis.v24i0.2519