Artificial Intelligence and Machine Learning as business tools: a framework for diagnosing value destruction potential

Ana Isabel Canhoto[a,b,*], Fintan Clear[a,c]

[a] Brunel University London, Kingston Lane, Uxbridge UB8 3PH, England; + 44.1895 274000

[b] ana.canhoto@brunel.ac.uk

[c] fintan.clear@brunel.ac.uk

[*] Corresponding author

Artificial Intelligence and Machine Learning as business tools: a framework for diagnosing value destruction potential

ABSTRACT

Artificial Intelligence (AI) and Machine Learning (ML) may save costs, and improve the efficiency of business processes. However, these technologies can also destroy business value, sometimes critically. The inability to identify how AI and ML may destroy value for businesses, and manage that risk, lead some managers to delay the adoption of these technologies, and, hence, prevents them from realizing the technologies' potential as business tools. This article proposes a new framework by which to map the components of an AI solution, and to identify and manage the value destruction potential of AI and ML for businesses. We show how the defining characteristics of AI and ML risk the integrity of the AI system's inputs, processes and outcomes. We, then, drawn on the concepts of value creation content and value creation process to conceptualize how these risks may hinder the process of value creation and actually result in value destruction. Finally, we illustrate the application of our framework with the example of the deployment of an AI powered chatbot in customer service, and discuss how to remedy the problems identified.

Keywords: Artificial Intelligence, Machine Learning, Value creation, Value destruction, Decision making, Technology adoption

1. AI AND ML USE IN THE BUSINESS CONTEXT

As we enter the fourth industrial revolution (Schwab, 2016), Artificial Intelligence (AI) and Machine Learning (ML) technologies are being used to automate business processes in more and more areas, from calculating optimal transport loads to shortlisting loan applicants without human input. These technologies promise to be more cost-effective than humans (Castelli, Manzoni & Popovič, 2017). However, they can also be problematic. For instance, automatic trading algorithms have created flash crashes in the US stock market (Varol, Ferrara, Davis, Menczer & Flammini, 2017); while Uber's self-driving vehicle hit and killed a pedestrian (Levin & Wong, 2018).

Surveys show that managers are delaying the adoption of AI and ML because they are unsure about how it can help their firms (Bughin, Chui, & McCarthy, 2017). Therefore, this paper aims to empower decision makers to identify the problems that may arise in their firms, so that they can manage the risks and be confident in their investment. Specifically, we propose a framework that considers the various components of an AI solution, their fundamental characteristics, and how these may result in the destruction of value for the business.

In the next section, we map the components of an AI solution, before we examine how the defining characteristics of AI and ML risk the integrity of the AI system's inputs, processes and outcomes. We, then, drawn on the concepts of value creation content and value creation process to conceptualize how these risks may actually result in value destruction for the firm. Finally, we illustrate the application of our framework with the example of the deployment of an AI powered chatbot in customer service. In our concluding remarks, we discuss how to remedy the problems identified.

2. COMPONENTS OF AN AI SOLUTION

We define AI as an assemblage of technological components which collect, process and act on data in ways that simulate human intelligence. For instance, AI solutions can apply rules, learn over time through the acquisition of new data and information, (referred to as ML), and adapt to changes in the environment (Russell & Norvig, 2016).

While there are numerous AI applications in an increasing range of industries, they all have three key components in common (Figure 1). The first of these components is the *input data*. According to the Merriam-Webster dictionary, the term 'data' refers to facts (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation. It differs from the term 'information', which, according to the same source, refers to knowledge obtained from investigation, study, or instruction. Data are so integral to the functioning of AI that, without them, AI has been described as mathematical fiction (Willson, 2017). AI can cope with large volumes of data, making it essential in the age of Big Data (Kietzmann, Paschen & Treen, 2018). Moreover, AI is increasingly able to use unstructured inputs such as images, speech or conversations, in addition to structured inputs like transaction data (Paschen, Pitt & Kietzmann, 2020). One common type of data used is historical data. For instance, Fraugster uses transaction data such as billing vs shipping address, and the type of IP connection used, to detect payment fraud (O'Hear, 2016). AI can also use data collected in real time, via physical sensors or by tracking online activity. For example, a retailer may use beacons to monitor loyal customers in the store, in combination with evidence that they are browsing a competitor's website via the store's

WiFi, to decide to offer them a discount. AI may also tap into the firm's knowledge databases, such as whether previous product recommendations were accepted or rejected.

The second key component is the *ML algorithm*, i.e., the computational procedure that processes the data inputs (Skiena, 2012). One type of ML is *supervised learning* where (human) experts give the computer training datasets with both the inputs and the correct outputs, for the algorithm to learn the patterns and to develop the rules to be applied to future instances of the same problem. For instance, AI can be trained to detect small cell variations in MRI scans to find early stage cancer (Tucker, 2018). The opposite approach is *unsupervised learning* where the computer is given a training dataset with inputs, but no labels. The ML's task is to find the best way of grouping the data points and how they may be related. This technique may be used to identify items that are purchased together. The final form is called *reinforcement learning*. The ML is given a training dataset plus a goal, and is left to find the best combination of actions to achieve that goal. It needs to be given criteria for judging alternative courses of actions (e.g., winning a game) and rewards for the actions that it takes (e.g., higher game score) (Mnih et al., 2013).

The third key AI component is the *output decision* resulting from the ML processing. At the lower end of the spectrum, AI may produce a single result, for instance a deception score (Elkins, Dunbar, Adame & Nunamaker Jr, 2013) which has no performative value until an analyst decides to act on it. Or the system may produce a selection of results for further action by human analysts, such as flagging content for the attention of moderators in online platforms. Finally, some AI systems have autonomy to act on the basis of the results of their analysis; for instance, a self-driving car can drive, steer or brake without human intervention (Goodal, 2016).

[Insert Figure 1 about here]

3. CHARACTERISTICS OF AI AND ML, AND THEIR IMPACT

The components described in the previous section work together because of certain characteristics which enable AI solutions, but may also impact negatively on each of the key components (see Table 1). The first such characteristic is *connectivity* between the various AI components. For instance, self-driving cars are connected to each other so that when one car makes a mistake, the learning can be quickly shared with the network. AI can also connect with external databases to use textual, visual, meta-data and other types of external data such as search engines (Bordino et al., 2012) or social media (Kalampokis, Tambouris & Tarabanis, 2013). As the business has no control over how external inputs were collected or labelled, it may be using data that are corrupted, incomplete or that mean something different from what the data label suggests. Connectivity also relies on the different parties being compatible with each other (e.g., the date needs to be entered in the same format across the system), though such standardization reduces AI's flexibility and limits its contextual richness (Alaimo & Kallinikos, 2017). Moreover, the need to use compatible programming languages may lead to particular algorithms being used for pragmatic reasons (Calvard, 2016) rather than because they are the best for the specific business problem (Skiena, 2012). Finally, poor outputs can spread broadly and quickly, increasing the scope and likelihood of mistakes. For example, bots that automatically aggregate news feeds' content can spread unverified information and rumors (Ferrara, Varol, Davis, Menczer & Flammini, 2014).

The second characteristic is *cognitive ability*. ML detects patterns in the input data, learns from mistakes and self-corrects. For instance, AlphaGo Zero has mastered the board game Go, simply by playing against itself over and over again (Silver et al., 2017). AI's cognitive ability has led to a move away from describing how consumers behave to predicting and, even, trying to influence that behavior, e.g., by personalizing the customer experience (Johar, Mookerjee & Sarkar, 2014). The quality of ML predictions is very difficult to assess prior to implementation and scaling (Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016), which presents risks. It is also difficult to assess whether the patterns identified through ML are true of the population at large, or only in terms of the data set available (Hudson, 2017). Moreover, ML can produce outputs that are not comprehensible to humans, and therefore, are impossible to correct or control, as when Facebook's AI negotiation bots developed their own, incomprehensible to humans, language (Lewis, Yarats, Dauphin, Parikh & Batra, 2017). AI's cognitive ability has also led to it being applied in areas that stretch ML's ability to convert complex features or ideas into binary formats. One example of over-simplification is the attempt to use AI to predict a person's sexual orientation based on facial features. The algorithm uses a binary definition of gender identity and sexual orientation, failing to reflect the variety of ways in which they can be defined, both physiologically and psychologically (Sharpe & Raj, 2017).

Third, *imperceptibility*. The vast majority of AI applications go unnoticed by users (Wilson & Daugherty, 2018), which can support acceptance of the technology, and satisfaction with the interaction. It can even improve user behavior as exemplified by Microsoft's chatbot, Tay, which had to be switched off following interactions with Twitter users who maliciously exploited a vulnerability in Tay's design (Lee, 2016). However, AI's imperceptibility also means that its use may go unchecked and unchallenged. This presents ethical and reputational threats, as data collection expands from explicit interactions between the firm and the customer, to include the customers' social life (Park, Huh, Oh & Pil, 2012) or even their home life, via personal wearables and other internet-enabled devices. There may also be fewer opportunities to correct mistakes and biases. Plus, it undermines the principles of choice and informed consent as illustrated by Google Duplex's AI voice assistant presentation (Solon, 2018b). In addition, the imperceptibility of AI makes it difficult to assess whether it is possible and secure to access the data needed. For instance, certain US law-enforcement agencies have been using AI to find criminals in a crowd. However, as the solution was developed by third parties, the agencies do not know what data the AI is using, what weight is given to different features, or what assumptions were made when defining the variables (Hudson, 2017). Firms may also be unable to access and update the underlying model, assumptions and data sources (Khan, Gadalla, Mitchell-Keller & Goldberg, 2016). Moreover, it has been noted that people act differently when they realize that they are interacting with AI (Lee, 2016). Without knowing whether the observations resulted from interactions with perceptible AI, managers cannot assess how representative of reality the data being modelled is.

[Insert Table 1 about here]

4. IDENTIFYING THE VALUE DESTRUCTION POTENTIAL OF AI

Value is a concept at the heart of the business literature (Jarvi, Kahkonen & Torvinen, 2018). It is a goal that the business tries to accomplish, either directly, through its own operations; or indirectly, by creating goods and services that the customer is willing to acquire (Bowman & Ambrosini, 2000). It’s for this reason that authors such as Urbinati, Bogers, Chiesa & Frattini (2019), and many others before them, have stated that the purpose of a business is to create value. However, value can also be destroyed, sometimes even resulting in the failure of firms that were once industry leaders (Rai & Tang, 2014). In this section, we follow Lepak, Smith and Taylor (2007) in considering both the value that can be created or destroyed (4.1), as well as how that happens (4.2).

4.1 What is value creation and destruction

In its simplest form, value creation is defined as the positive contribution to the utility of the target user, and occurs any time the benefit of a business action (for instance, changes in a task, or the development of a new product) outweigh its cost (Porter 1985). Value is subjective and specific, in the sense that it is judged by the target user, in terms of its appropriateness to the task at hand, and relative to the closest alternative (Bowman & Ambrosini, 2000). The contextual nature of value means that we need to evaluate AI and ML in light of the specific tasks that are meant to be performed by the technology, and relative to the relevant alternative investment.

Conversely, value destruction occurs when there is a perceived reduction in utility, either because the target user does not perceive any net benefits for the task at hand. For instance, stakeholders may disagree on whether the outcome of a project is positive or negative, or even on which criteria to use (Willumsen, Oehmen, Stingl & Geraldi, 2019).

Value creation may take the form of novel, efficient or complementary solutions (Rai & Tang, 2014). Novelty occurs when new components are connected to each other, or in new ways (Amit & Zott, 2001). Efficiency occurs by streamlining activities (Zott & Amit, 2007), and complementarity by integrating assets with network effects (Rai & Tang, 2014).

The main form of assessing the performance of AI and ML in business settings is its cost-efficiency. AI solutions are said to be cheaper, faster and less prone to mistakes than humans (Castelli et al, 2017), particularly when applied to mechanical and analytical tasks (Huang & Rust, 2018). For instance, self-driving cars may be better than humans at avoiding road collisions (Goodal, 2016). Though, AI and, particularly, ML are also valued for their ability to produce novel outcomes, such as finding previously unknown patterns in the datasets available (Kietzmann et al, 2018); or new ways of solving a problem (Silver et al., 2017). In addition, the connectivity aspect of AI and ML enable complementarity among different nodes in a network, such as individual vehicles in a self-driving fleet.

However, if the cost of achieving these benefits is narrowly defined, it may underestimate costs such as reputational damage. For instance, the public's concern with the ethical problems associated with the decisions embedded in self-driving cars' algorithms, such as whether to protect the life of the vehicle's occupants or the by-standers' "*risks marginalizing the entire field*" (Goodall, 2016, p. 810). Cost calculation may also fail to account for trade-offs such as calculation speed vs. degree of confidence in the calculation's results (Cormen, Leiserson, Rivest & Stein, 2001), or accuracy vs. interpretability of the algorithm (Lee & Shin, 2020). Trade-offs can also occur over time. For instance, accuracy can be increased if the business is prepared to allow for mistakes in the short-term, or to invest in quality checkers to train the ML (Solon, 2018a). Another issue to consider is the business's starting point. Analytical capabilities and big data handling skills vary significantly across firms (Merendino et al., 2018), which means that different firms will face different hurdles when deploying AI and ML.

4.2 How value is created or destroyed

The value creation process refers to the series of actions that result in the production of a net positive outcome. Conversely, the value destruction process is one that produces a negative outcome (Jarvi et al, 2018). The value creation literature has paid little attention to the causes or antecedents of value destruction (Prior & Marcos-Cuevas, 2016). Yet, it is vital for managers to understand the reasons for value destruction according to the specific phase in which they occur (Prior & Marcos-Cuevas, 2016). This way, managers can identify the pitfalls and adopt preventive or remedial action (Jarvi et al, 2018).

The normative literature states various prescriptive guidelines and best practice to be followed in order to ensure a successful outcome (Willumsen et al, 2019). From this perspective, value destruction may occur if the business does not follow the guidelines, and there are failures in the interaction process (Prior & Marcos-Cuevas, 2016). However, Troilo and colleagues (2017) argue that there is a lack of strategic frameworks that explain how value is created in the digital context and from big data, including for AI solutions.

Value destruction may also occur if the participants do not possess certain critical resources (Jarvi et al, 2018). Of particular relevance for AI solutions is the lack of access to key data and information (Vafeas, Hughes & Hilton, 2016), and IT assets (Benaroch & Chernobai, 2017). Lack of suitable IT resources has been shown to actively "*destroy value in a firm rather than simply fail to add any*" (Arend, 2003, p. 280). Goldstein and colleagues (2011) go in so far as to assert that the lack of functional IT resources is more harmful to a firm's value creation efforts than data protection failures.

In addition, value creation requires that firms embrace change (Jarvi et al, 2018) and adapt their behavior accordingly (Homburg, Jozić, & Kuehnl, 2017). Digital technologies, in particular, require firms to change their behavioral models and how they interact with their stakeholders (Jarvi et al, 2018). Yet, research (e.g., Merendino et al, 2018) shows that many organizations struggle to adapt their strategic decision-making processes and procedures to reflect the changes caused by Big Data, AI and other such technologies.

5. IMPLEMENTING THE FRAMEWORK

We now discuss how the theoretical concepts previously presented may be used as a diagnostic tool (as per Figure 2) to support managers in diagnosing when deploying AI solutions may result in value destruction for their firms. We illustrate the managerial application of our framework, with the example of deploying an AI chatbot to handle customer complaints on Twitter.

5.1 Mapping out the components of the solution

We start by identifying the various components of the chatbot solution, and the risks presented by their connectivity, cognitive ability and imperceptibility (Table 2). Starting with the *input data*, the business needs to have a channel to collect the comments from customers, in real time. In this case, the channel used is Twitter, which is a channel external to the firm, with its own policies and practices, and whose operations (e.g., website maintenance) are beyond the firm's influence. While the customer chooses to interact with the firm via Twitter, they may not be aware that they will be interacting with a chatbot, or that the data provided is also being collected and analyzed by Twitter itself.

The chatbot also needs access to be connected to FAQs databases, inventory and other sources of data, as well as the company's customer support team (Wilson & Daugherty, 2018). In addition, the bot needs to access a database of historical customer data such as past interactions, as well as customer information such as the customer's lifetime value or propensity to churn, to be able to personalize the answers (Kietzmann *et al*., 2018).

In turn, the ML *algorithm* will need to process Twitter's free-form text. It must be capable of handling natural language processing in order to analyze and respond to the customers' comments. The algorithm should also be able to identify the desired outcome, understand whether the customer is getting upset and identify the course of action most likely to meet the customer's needs.

Finally, the chatbot needs to perform an action. There are four types of task possible (Huang & Rust, 2018): a mechanical task, such as delivering a scripted response based on key words used by the customer; an analytical task, such as being able to reach a conclusion about the type of problem faced by the customer; an intuitive task, such as understanding why the customer is complaining; or an empathetic task, such as trying to calm down an upset customer. The former tasks are easier to perform than the latter ones, even for very powerful AI solutions.

The task can be performed autonomously (e.g., provide delivery information) or through a member of staff (e.g., approval for a refund note to be issued), which requires connectivity. If interacting directly with the customer, the bot requires natural language generation abilities, in order to produce a reply that is intelligible to non-experts and adapted to the circumstances of the complaint.

[Insert Table 2 about here]

5.2 Predicting how value may be destroyed

Businesses have long adopted some form of automation in complaint handling (e.g., via FAQ pages on websites), because many complaints are quite common and have relatively easy-to-mechanize solutions. Examples include delayed deliveries, the need to return or exchange items, and requests for compensation.

In this scenario, the chatbot will be interacting with customers on Twitter, which means that the business can use an API (Application Programming Interface) or processing application to automate collection and analysis of the tweets, profile data, and meta-data (e.g., location). However, if this is not a channel favored by the customers, then Twitter is not a relevant resource for complaints handling.

Given that up to 15% of current Twitter accounts are controlled by malicious bots (Varol *et al*., 2017), there is a risk that the chatbot may not be interacting with a real customer, which would waste the firm's processing resources. It could also lead to misinformation being fed into the ML algorithm, and perhaps embarrassment if the exchanges resulted in comical or offensive replies, as illustrated in the Tay's case previously mentioned.

Once it is clear that the chatbot is interacting with a person, it has to decode what the customer is saying, both explicitly and implicitly, and detect sentiment. The chatbot may be unable to collect all data formats shared on a platform. For instance, even though technology is now capable of collecting unstructured data, many businesses do not use such technology due to limited budgets or incompatibility with legacy systems. Alternatively, it may be unable to use unstructured data that it has collected – e.g., images with low resolution (Soton, 2018a); or draw on all data sources available, because doing so would require processing power beyond existing capacity (Agarwal, 2014). Moreover, while the chatbot may be programmed to detect common sentiment features indicating the valence (e.g., through certain keywords) and intensity (e.g., use of capital letters and exclamation points), it is likely to struggle with humor and irony (Canhoto & Padmanabhan, 2015). Bots also struggle with spelling mistakes and multiple languages, which is problematic for companies with presence in countries with more than one official language (e.g., Canada).

In terms of the algorithm, it is crucial that it uses a technique that matches the type of problem. Hence, the business needs to know and understand what the algorithm does and how it reaches conclusions. This is likely to be a challenge for two reasons. First, many businesses use algorithms developed by third parties who do not disclose what they see as proprietary information. Second, many senior managers may not have the necessary technical skills or even the type of (non-linear) thinking required (Merendino et al, 2018).

The algorithms will need constant updating (Khan *et al*., 2016), for instance, to reflect a change in regulations, a new product line, or a recent promotional activity. Otherwise, the algorithms lose integrity and contextual relevance to address the complaint. Moreover, the algorithms need to have both mathematical insight and rules, as well as assumptions about the world (ibid.). If the programmers have made an incorrect assumption (e.g., regarding words that have different meanings depending on context), this can lead to unsatisfactory results and the destruction of value.

It is also possible that the customer does not realize that it is interacting with a bot, and becomes frustrated with the exchange when, for instance, the bot asks a question that does not follow meaningfully from what has just been said. Also, the maintenance of internal databases often requires collaboration from staff for data input. For instance, FAQs databases might require staff to record all questions, including unusual ones. If there are limited financial resources or the employees are unwilling to cooperate, the resulting database will be incomplete. Another question concerns how much historical data the bot has access to, and how representative the database of possible solutions may be.

If the business opts for supervised or reinforced learning, it may encounter problems in situations where there is no simple set of rules to link the variables or to rank the outcomes. For instance, many malicious bot accounts adopt characteristics that hinder their detection (Varol *et al*., 2017). On the other hand, unsupervised learning can create self-reinforcing feedback loops, quickly becoming so complex that even the people who created them can no longer explain how they work (Hudson, 2017).

5.3 Assessing what value could be destroyed

In recent years, there has been an increase in the use of AI and ML to handle online customer complaints. These solutions allow for real-time, personalized replies (Kietzmann *et al.*, 2018). They also reduce the customers' cost of complaining, which may incentivize customers to voice their dissatisfaction directly to the firm (Istanbulluoglu, Leek & Szmigin, 2017). However, customer complaints are also a critical point for customer satisfaction and recovery following a service failure (Istanbulluoglu et al, 2017).

The first question that arises in terms of the value of the output, concerns how a "good outcome" is defined. The interests of the business and the customer are likely to diverge (Dawar, 2018), and while a human customer service assistant may be able to strike the best balance between the two, this is unlikely to be the case with a chatbot. AI deals best with mechanical tasks and very well with analytical ones, but struggles with intuitive or empathetic tasks (Huang & Rust, 2018). This is a problem in complaint management, where intuition and empathy are key, in order to understand the type of outcome sought by the client or to dissolve tension (Istanbulluoglu *et al*., 2017).

Firms may be tempted to fully automate their conversations with customers in order to maximise cost savings. However, these tend to lead to unsatisfactory results. For instance, instead of solving the problem quickly, the chatbot may end up creating confusion or adding unnecessary delays to the interaction. The bot may also destroy value by producing a response that is not aligned with the brand image or persona (CCW, 2017). Instead, it may be better to experiment with different combinations of staff and AI, as human agents are likely to be better at adapting their styles to different audiences (CCW, 2017). The AI can produce data visualizations that help analysts identify particular patterns, and provide answers to customers (Khan *et al*., 2016). However, these too need to be adapted to the person using the visualizations, and be complemented with training.

6. CONCLUDING REMARKS

Having presented a tool to diagnose how AI solutions can create problems for the businesses deploying them, we now reflect on the issues likely to be faced when solving those problems. First, amidst the plethora of issues requiring the managers' attention, they need to be able to identify those they should devote their attention to (Davenport & Beck, 2002). For that, managers have to quantify the potential for value destruction associated with each of the components of the AI solution. Upon consultation with key stakeholders in the organization such as database managers or brand managers, the impact of each problem detected via application of our framework can be ranked using a Likert scale, thus producing a ranking of events that would lead to the greatest destruction of value for the business, as well as the likelihood of said event occurring. Then, the manager can produce a visualization of the source and impact of each risk, to help communicate the magnitude of the potential problems to others in the firm (Lowy & Hood, 2004). For instance, an untested ML model represents a potential weakness in terms of the processing algorithms' cognitive ability. If the AI powered solution is not connected to other components and is used with a small group of customers only, it does not represent a big risk. However, if the solution is being deployed quickly, or if it is providing advice in regulated industries such as financial services, small mistakes can easily escalate to big problems.

Second, there are the costs of actually preventing or addressing the problems identified. AI and ML can represent a heavy investment, both initially and in terms of on-going maintenance. Even the simplest AI solution requires heavy initial investment in training (Soton, 2018a). It will also require processing power, access to various databases, and regular updating, all of which are costly. AI and ML also require specialist skills which most businesses are lacking. Some firms find that recruiting talent with those skills is difficult and expensive; while others opt for outsourcing (Merendino *et al*., 2018).

Finally, using AI and ML requires multiple decisions and difficult choices about value and values (Hudson, 2017). It requires the business to face ingrained biases in the way it operates, and which negatively impact the quality of the input data, training datasets and algorithms. The business needs to decide what type of accuracy is the most important, and whether it prefers to incur a false negative error, or a false positive one. Likewise, the business needs to define fairness, and decide whom it is most concerned with treating fairly. Moreover, the outcomes produced through AI may be highly consequential for the firm and its customers. Predicting someone's sexuality may sound innocuous when it comes to personalizing an advert, but could result in one group not being given the same opportunities as others, which is a form of discrimination. It can also increase some customers' social and economic vulnerability, or, in some places, put them in life-threatening situations (Sharpe & Raj, 2017). Hence, deploying AI requires businesses to consider the consequences of their actions beyond first order effects.

In summary, if businesses wish to benefit from using AI and ML tools, then it is clear that a sophisticated understanding of the tools and careful analysis of the risks will be required, as well as some initial investment, in order to avoid inadvertent value destruction. Use of the framework presented in this paper should spur managers to look beyond the type of algorithm used and incomplete cost-benefit calculations, thus ensuring that they avoid some common pitfalls and can truly create value for their businesses.

REFERENCES

Agarwal, R. (2014). Editorial - Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, 25, 443-448.

Alaimo, C. & Kallinikos, J., (2017). Computing the everyday: social media as data platforms. *Information Society*, 33(4), 175-191.

Amit, R. & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, 22(6/7), 493–520.

Arend, R. J. (2003). Revisiting the Logical and Research Considerations of Competitive Advantage. *Strategic Management Journal*, 24(3), 279-284.

Benaroch, M. & Chernobai, A. (2017). Operational IT Failures, IT Value Destruction, and Board-Level IT Governance Changes. *MIS Quarterly*, 41(3), 729-762.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A. & Weber, I. (2012). Web Search Queries Can Predict Stock Market Volumes, *PLoS ONE*, 7(7).

Bowman, C. & Ambrosini, V. (2000). Value creation versus value capture: towards a coherent definition of value in strategy. *British Journal of Management*, 11 (1), 1–15.

Bughin, J., Chui, M. & McCarthy, B. (2017). *What every CEO needs to know to succeed with AI*. McKinsey. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/digital-blog/what-every-ceo-needs-to-know-to-succeed-with-ai

Calvard, T. S. (2016). Big data, organizational learning, and sensemaking: Theorizing interpretive challenges under conditions of dynamic complexity. *Management Learning*, 47(1), 65-82.

Canhoto, A. I. & Padmanabhan, Y. (2015). We (don't) know how you feel – A comparative study of automated vs. manual analysis of social media conversations. *Journal of Marketing Management*, 31(9-10), 1141-1157.

Castelli, M., Manzoni, L. & Popovič, A. (2016). An Artificial Intelligence System to Predict Quality of Service in Banking Organizations. *Computational Intelligence and Neuroscience*, Article ID 9139380.

CCW. (2017). *Chatbots, Everything you need to know*. *CCW*. https://www.customercontactweek.com/downloads/chatbots-everything-you-need-to-know?-ty-m

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (200)1. *Introduction to Algorithms*. London: McGraw-Hill, 2nd edition.

Davenport, T. H. & Beck, J. C. (2002). *Attention Economy: Understanding the New Currency of Business.* Cambridge MA: Harvard Business Review Press.

Dawar, N. (2018). Marketing in the Age of Alexa. *Harvard Business Review*, 96(3), 80-86.

Elkins, A. C., Dunbar, N. E., Adame, B. & Nunamaker Jr, J. F. (2013). Are Users Threatened by Credibility Assessment Systems? *Journal of Management Information Systems*. 29(4), 249-261.

Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. (2014). The Rise of Social Bots. *Communications of ACM*, 59(7), 96-104.

Goldstein, J., Chernobai, A., & Benaroch, M. (2011). Event Study Analysis of the Economic Impact of IT Operational Risk and its Subcategories. *Journal of the Association of Information Systems*, 12(9), 606-631.

Goodall, N. J. (2016). Away from Trolley Problems and Toward Risk Management. *Applied Artificial Intelligence*, 30(8), 810-821.

Homburg, C., Jozić, D. & Kuehnl, C. (2017). Customer experience management: toward implementing an evolving marketing concept. *Journal of the Academy of Marketing Science*, 45(3), 377-401.

Huang, M-H & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155-172.

Hudson, L. (2017). Technology Is Biased Too. How Do We Fix It? *FiveThirtyEight*. https://fivethirtyeight.com/features/technology-is-biased-too-how-do-we-fix-it/amp/

Istanbulluoglu, D., Leek, S. & Szmigin, I. T. (2017). Beyond exit and voice: developing an integrated taxonomy of consumer complaining behaviour. *European Journal of Marketing*, 51(5-6), 1109-1128.

Järvi, H., Kähkönen, A-K. & Torvinen, H. (2018). When value co-creation fails: Reasons that lead to value co-destruction. *Scandinavian Journal of Management*, 34(1), 63-77.

Johar, M., Mookerjee, V. & Sarkar, S. (2014). Selling vs. Profiling: Optimizing the Offer Set in Web-Based Personalization. *Information Systems Research*, 25(2), 285-306.

Kalampokis, E. Tambouris, E. & Tarabanis, K. (2013). Understanding the predictive power of social media. *Internet Research*, 23(5), 544-559.

Khan, I., Gadalla, C., Mitchell-Keller, L. & Goldberg, M. S. (2016). Algorithms: The New Means of Production. *D!gitalist*. https://www.digitalistmag.com/executive-research/algorithms-the-new-means-of-production

Kietzmann, J.H., Paschen, J. & Treen, E. (2018). Artificial Intelligence in Advertising: How Marketers Can Leverage Artificial Intelligence Along the Consumer Journey. *Journal of Advertising Research*, 58(3), 263-267.

Lee, P. (2016). *Learning from Tay's introduction*. Microsoft. https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.00000gjdpwwcfcus11t6oo6dw79gw

Lee, I. & Shin, Y. J. (2020). Machine Learning for Enterprises: Applications, Algorithm Selection, and Challenges. *Business Horizons*, 63(1).

Lepak, D. P., Smith, K.G., & Taylor, M.S (2007), Introduction to Special Topic Forum: Value Creation and Value Capture: A Multilevel Perspective, *The Academy of Management Review*, 32(1), 180-194

Levin, S. & Wong, J. C. (2018). Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. *The Guardian*. https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe

Lewis, M., Yarats, D., Dauphin, Y. N., Parikh, D. & Batra, D. (2017). *Deal or no deal? Training AI bots to negotiate*. Facebook Code. https://code.fb.com/ml-applications/deal-or-no-deal-training-ai-bots-to-negotiate/

Lowy, A. & Hood, P. (2004). *The Power of the 2x2 Matrix: Using 2x2 Thinking to Solve Business Problem and Make Better Decisions*. London: Jossey-Bass

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L. & Canhoto, A. I. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, 93(December), 67-78.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. & Riedmiller, M. (2013). Playing Atari with Deep Reinforcement Learning. *NIPS Deep Learning Workshop*. Cornell University.

O'Hear, S. (2016). Fraugster, a startup that uses AI to detect payment fraud, raises $5M, *TechCrunch*. https://techcrunch.com/2017/01/16/fraugster/

Park, S-H., Huh, S-Y., Oh, W. & Pil, S. (2012). A Social Network-Based Inference Model for Validating Customer Profile Data. *MIS Quarterly*, 36(4), 1217-1237.

Paschen, J., Pitt, C. & Kietzmann, J. (2020). Artificial Intelligence: Building Blocks and an Innovation Typology. *Business Horizons*, 63(1).

Porter, M.E. (1985). *Competitive Advantage. Creating and Sustaining Superior Performance*. Free Press: New York

Prior, D. D. & Marcos-Cuevas, J. (2016). Value co-destruction in interfirm relationships: The impact of actor engagement styles. Marketing Theory, 16(4), 533-552.

Rai, A. & Tang, Z. (2014). Information Technology-Enabled Business Models: A Conceptual Framework and a Coevolution Perspective for Future Research. *Information Systems Research*, 25(1), 1-14.

Russell, S. J. & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. London: Pearson, 3rd edition.

Schwab, K. (2016). *The Fourth Industrial Revolution*. Geneva: World Economic Forum.

Sharpe, A. & Raj, S. (2017). Using AI to determine queer sexuality is misconceived and dangerous. *The Conversation*. https://theconversation.com/using-ai-to-determine-queer-sexuality-is-misconceived-and-dangerous-83931?utm_source=twitter&utm_medium=twitterbutton

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(October), 354-359.

Skiena, S. S. (2012). *The algorithm design manual*. London: Springer.

Solon, O. (2018a). The rise of 'pseudo-AI': how tech firms quietly use humans to do bots' work. *The Guardian*. https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies?CMP=Share_AndroidApp_Direct_Message

Solon, O. (2018b). Google's robot assistant now makes eerily lifelike phone calls for you. *The Guardian*. https://www.theguardian.com/technology/2018/may/08/google-duplex-assistant-phone-calls-robot-human

Troilo, G., de Luca, L. M. & Guenzi, P. (2017). Linking Data-Rich Environments with Service Innovation in Incumbent Firms: A Conceptual Framework and Research Propositions. *Journal of Product Innovation Management*, 34(5), 617-639.

Tucker, I. (2018). AI cancer detectors. *The Guardian*. https://www.theguardian.com/technology/2018/jun/10/artificial-intelligence-cancer-detectors-the-five

Urbinati, A., Bogers, M., Chiesa, V. & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation*, 84-85(June-July), 21-36.

Vafeas, M., Hughes, T., & Hilton, T. (2016). Antecedents to value diminution: A dyadic perspective. *Marketing Theory*, 16(4), 469-491.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F. & Flammini A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. *ICWSM'17*.

Willson, M. (2017). Algorithms (and the) everyday. *Information, Communication & Society*, 20(1), 137-150.

Willumsen, P., Oehmen, J., Stingl, V. & Geraldi, J. (2019). Value creation through project risk management. *International Journal of Project Management*, 37(5), 731-749.

Wilson, H., J. & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review*, 96(4), 114-123.

Zott, C. & Amit, R. (2007). Business model design and the performance of entrepreneurial firms. *Organization Science*, 18(2), 181–199

Table 1. The impact of connectivity, cognitive ability and imperceptibility

Component	Connectivity	Cognitive ability	Imperceptibility
Input data	Use of external data over which the firm has limited quality control	Dataset may be unsuitable for predictive profiling	User unable to provide informed consent; data may not be representative
Processing algorithm	Trade-off between standardization and compatibility vs. fit and flexibility	Formulae oversimplify complex phenomena	No ability to access, assess and update model
Output decision	Mistakes and poor outputs can go viral	Difficulty in verifying quality of predictions, or even understand ML outputs	Impossible to check, challenge or correct outcomes

Table 2. Assessment of the risks presented by a chatbot's connectivity, cognitive ability and imperceptibility

Component	Connectivity	Cognitive ability	Imperceptibility
Input data <u>*Description:*</u> *Free form text; Responses and solutions; Past transactions; Customer data*	Relies on access to real time data from external source; Requires links to various internal databases	Needs to process different types of internal and external data, including unstructured data	User unable to provide informed consent
Processing algorithm <u>*Description:*</u> *Natural language processing (NLP); Sentiment analysis; Result ranking*	To correctly assess sentiment, needs access to contextual information	NLP offers flexibility but increases likelihood of error, and processing capability	May be unable to assess result ranking
Output decision <u>*Description:*</u> *Action; Natural language generation (NLG); Staff intervention*	Requires seamless integration between chatbot and staff	Sophisticated cognitive ability required to respond to emotions	If using NLG, user may be unable to challenge or correct outcomes

Figure 1. Key components of an AI solution

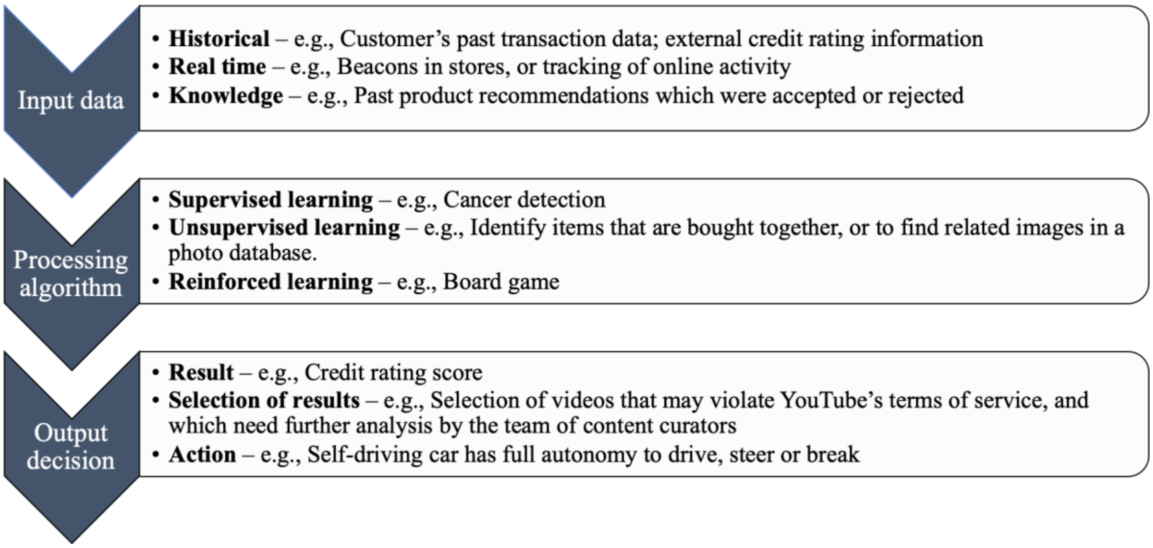

Figure 2. Diagnosing the value destruction potential of a business AI solution

Step	Diagnostic items
1. Identify risks created by AI characteristics, for each component of the solution	1. Risks created by connectivity, given the type(s) of input data (historical, real time, or knowledge) used 2. Risks created by cognitive ability, given the type(s) of input data (historical, real time, or knowledge) used 3. Risks created by imperceptibility, given the type(s) of input data (historical, real time, or knowledge) used 4. Risks created by connectivity, given the type(s) of algorithm (supervised, unsupervised or reinforced learning) used 5. Risks created by cognitive ability, given the type(s) of algorithm (supervised, unsupervised or reinforced learning) used 6. Risks created by imperceptibility, given the type(s) of algorithm (supervised, unsupervised or reinforced learning) used 7. Risks created by connectivity, given the type(s) of output (result, selection of results or action) produced 8. Risks created by cognitive ability, given the type(s) of output (result, selection of results or action) produced 9. Risks created by imperceptibility, given the type(s) of output (result, selection of results or action) produced
2. Analyze how the risks identified in step 1 may destroy business value	1. Relevant guidelines for handling the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of input data used 2. Relevant guidelines for handling the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of algorithm used 3. Relevant guidelines for handling the connectivity, cognitive ability and imperceptibility risks associated with the type(s) output produced 4. Resources required (including data, information, and IT assets) to handle the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of input data used 5. Resources required (including data, information, and IT assets) to handle the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of algorithm used 6. Resources required (including data, information, and IT assets) to handle the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of output produced 7. Behavioral changes required to handle the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of input data used 8. Behavioral changes required to handle the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of algorithm used 9. Behavioral changes required to handle the connectivity, cognitive ability and imperceptibility risks associated with the type(s) of output produced
3. Recognize what type of business value may be destroyed by failing to follow the process principles described in step 2.	1. Impact on ability to innovate 2. Impact on cost-efficiency 3. Impact on complementarity