Business Horizons (xxxx) xxx, xxx

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY

www.journals.elsevier.com/business-horizons

Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential

Ana Isabel Canhoto*, Fintan Clear

Brunel University London, Kingston Lane, Uxbridge UB8 3PH, U.K.

KEYWORDS
Artificial intelligence;
Machine learning;
Value creation;
Value destruction;
Decision making;
Technology adoption;
Enterprise value;
Chatbot

Abstract Artificial intelligence (AI) and machine learning (ML) may save money and improve the efficiency of business processes, but these technologies can also destroy business value, sometimes with grave consequences. The inability to identify and manage that risk can lead some managers to delay the adoption of these technologies and thus prevent them from realizing their potential. 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 can threaten the integrity of the AI system's inputs, processes, and outcomes. We then draw from the concepts of value-creation content and value-creation process to show how these risks may hinder value creation or even result in value destruction. Finally, we illustrate the application of our framework with an example of the deployment of an AI-powered chatbot in customer service, and we discuss how to remedy the problems that arise.

1. Artificial intelligence and machine learning in business

As we enter the fourth industrial revolution (Schwab, 2016), artificial intelligence (AI) and machine-learning (ML) technologies are driving business automation in more and more areas, from calculating optimal transport loads to short-listing loan applicants without human input. These technologies promise to be more cost-effective than humans (Castelli, Manzoni, & Popovič, 2016), but they can also be problematic. For instance, automatic trading algorithms have created flash crashes in the U.S. stock market (Varol, Ferrara, Davis, Menczer, & Flammini, 2017), and one of Uber's self-driving vehicles hit and killed a pedestrian (Levin & Wong, 2018).

* Corresponding author
E-mail addresses: ana.canhoto@brunel.ac.uk (A.I. Canhoto), fintan.clear@brunel.ac.uk (F. Clear)

https://doi.org/10.1016/j.bushor.2019.11.003

Surveys show that managers are delaying the adoption of AI and ML because they are unsure how these technologies can help their firms (Bughin, Chui, & McCarthy, 2017). This article aims to empower decision makers to identify problems so that they can manage risks and be confident in their investments. 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 and examine how the defining characteristics of AI and ML can threaten the integrity of the AI system's inputs, processes, and outcomes. We then draw from the concepts of value-creation content and value-creation process to show how these risks may result in value destruction for the firm. Finally, we illustrate the application of our framework with an example of the deployment of an AI-powered chatbot in customer service. In our concluding remarks, we discuss how to remedy the problems identified.

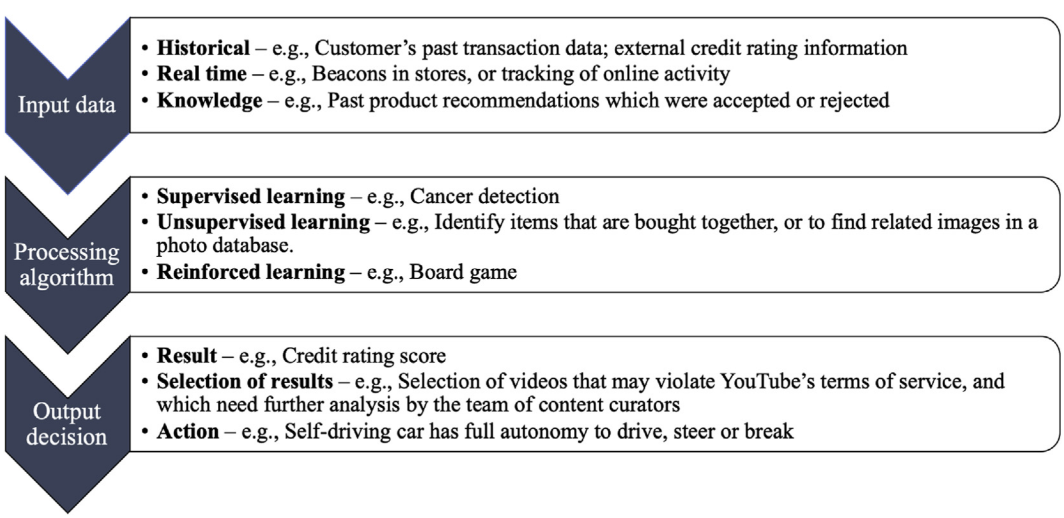
2. Components of an AI solution

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

While AI applications see use across an ever-increasing range of industries, they all have three components in common (see Figure 1). The first of these components is input data. Input 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 increasingly important in the dawning 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). Many companies use historical data in their AI applications. For instance, Fraugster uses transaction data, including billing and shipping addresses and IP connection type, to detect payment fraud (O'Hear, 2017). AI can also use data collected in real time, either via physical sensors or by tracking online activity. For example, a retailer's AI application may use beacons that monitor shoppers in the store, in combination with evidence that they are browsing a competitor's website via the store's wi-fi, to decide to offer them a discount. AI may also tap into the firm's databases to check whether those same shoppers accepted or rejected previous product recommendations.

The second key AI component is the ML algorithm, which is the computational procedure that processes the data inputs (Skiena, 2012). There are three types of ML algorithms: supervised, unsupervised, and reinforcement. In supervised ML, human experts give the computer training data sets with both the inputs and the correct outputs so the algorithm can learn the patterns and develop rules to be applied to future

Figure 1. Key components of an AI solution



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). By contrast, in unsupervised learning, the computer is given a training data set with inputs but no labels. The algorithm's task is to find the best way of grouping the data points and to establish how they may be related. This technique may be used to identify items that are purchased together, for example, which can inform marketing and sales strategies. The final form of ML is called reinforcement learning. Here the algorithm is given a training data set plus a goal, and it then must find the best combination of actions to achieve that goal. For this to work, 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 scores; Mnih et al., 2013).

The third key AI component is the output decision resulting from the ML process. At the lower end of the spectrum, AI may produce a single result—for instance, a deception score (Elkins, Dunbar, Adame, & Nunamaker, 2013) that 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 by 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 (Goodall, 2016).

3. Characteristics and effects of AI and ML

The key components described in the previous section work together because of certain characteristics that enable AI solutions but may also degrade or limit those same 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, metadata, and other types of external data, including search engines (Bordino et al., 2012) or social media (Kalampokis, Tambouris, & Tarabanis, 2013). Connectivity can impinge upon AI components in many ways. As a business may have no control over how external inputs are collected or labeled, its data could be corrupted, incomplete, or misleading. Connectivity also depends on the different parties being compatible with each other (e.g., dates need 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 cause a business to choose particular algorithms for pragmatic reasons (Calvard, 2016) rather than because they are the best for the specific problem at hand (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 of AI is its cognitive ability. ML algorithms detect patterns in the input data, learn from mistakes, and self-correct. 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 caused a shift from merely describing how consumers behave to predicting and even

Table 1. The effects 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	Formulas oversimplify complex phenomena	No ability to access, assess, or update model
Output decision	Mistakes and poor outputs can go viral	Difficulty in verifying quality of predictions, or even understanding ML outputs	Impossible to check, challenge, or correct outcomes

trying to influence that behavior (e.g., by personalizing the customer experience; Johar, Mookerjee, & Sarkar, 2014). But as with connectivity, AI's cognitive ability presents many challenges. 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 whether they describe only the data set available (Hudson, 2017). Moreover, ML can produce outputs that are incomprehensible to humans and therefore impossible to correct or control, as when Facebook's AI negotiation bots developed their own incomprehensible language (Lewis, Yarats, Dauphin, Parikh, & Batra, 2017). And there are limits to ML's ability to convert complex features or ideas into binary formats. For example, efforts to use AI to predict a person's sexual orientation according to their facial features have resulted in oversimplification. In one case, the algorithm used binary definitions of gender identity and sexual orientation, thus failing to reflect the variety of ways in which they can be defined, both physiologically and psychologically (Sharpe & Raj, 2017).

AI's third defining characteristic is its imperceptibility. The vast majority of AI applications go unnoticed by users (Wilson & Daugherty, 2018), which can aid users' acceptance of the technology. Imperceptibility can even improve user behavior, as some people are prone to misbehave when they realize they're interacting with AI, 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). But AI's imperceptibility also means that its use may go unchecked and unchallenged. This presents ethical and reputational threats, as data collection has expanded from explicit interactions between the firm and its customers to include the customers' social lives (Park, Huh, Oh, & Pil, 2012) and even their home lives via personal wearables and other internet-enabled devices. Imperceptible interactions may also yield less feedback and thus fewer opportunities to correct mistakes and biases. And imperceptibility undermines the principles of choice and informed consent, as illustrated by Google's Duplex AI voice-assistant presentation (Solon, 2018b). In addition, the imperceptibility of AI makes it difficult to assess the security of the data needed. For instance, certain U.S. law-enforcement agencies have been using AI to spot criminals in a crowd, but because the solution was developed by third parties, the agencies do not know what data the AI is using, how different features are weighted, 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 users themselves know they are interacting with AI, managers cannot assess how representative of reality their data may be.

4. Identifying the value-destruction potential of AI

Value is a concept at the heart of the business literature (Järvi, Kähkönen, & Torvinen, 2018). Businesses create value either directly, through their own operations, or indirectly, by creating goods and services that their customers are willing to acquire (Bowman & Ambrosini, 2000). Urbinati, Bogers, Chiesa, and Frattini (2019), and many others before them, have stated that the purpose of a business is to create value. But value can also be destroyed, which sometimes results even 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 value creation and destruction, as well as how those processes happen.

4.1. What are value creation and destruction?

In its simplest form, value creation is defined as the positive contribution to the utility of the target user. It occurs any time the benefits of a business action—for instance, the development of a new product—outweigh its costs (Porter, 1985). Value is subjective and specific: subjective in that it is judged by the target user, and specific in terms of its appropriateness to the task at hand and its relative benefits and costs compared 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 at hand and relative to any possible alternative investment.

Conversely, value destruction occurs when the target user perceives a reduction in utility. In some cases, stakeholders may disagree on whether the outcome of a project is positive or negative, or even on which criteria should decide this (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 when existing ones are connected in new ways (Amit & Zott, 2001). Efficiency occurs through streamlining activities (Zott & Amit, 2007), and complementarity through integrating assets with network effects (Rai & Tang, 2014).

The main way to assess the performance of AI and ML in business settings is by calculating their cost efficiency. AI solutions are said to be cheaper, faster, and less prone to mistakes than humans (Castelli et al., 2016), 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 (Goodall, 2016). But AI and especially ML are also valued for their ability to produce novel outcomes, such as finding previously unknown patterns in the available data sets (Kietzmann et al., 2018) or new ways of solving a problem (Silver et al., 2017). In addition, the connectivity aspects of AI and ML enable complementarity among different nodes in a network, such as individual vehicles in a self-driving fleet.

Businesses should beware of underestimating the potential costs of AI and ML, including the potential for 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 a vehicle's occupants at the expense of bystanders, "risks marginalizing the entire field" (Goodall, 2016, p. 810). Cost calculation may also fail to account for trade-offs such as calculation speed versus confidence (Cormen, Leiserson, Rivest, & Stein, 2001) or accuracy versus interpretability of the algorithm (Lee & Shin, 2020). These trade-offs can be revised over time. For instance, a business can work to increase its algorithm's long-term accuracy even as it allows for some mistakes in the short-term, which it can mitigate by investing in quality checkers to train the algorithm (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 and destroyed

The value-creation process is the series of actions that results in the production of a net positive outcome. Conversely, the value-destruction process is one that produces a negative outcome (Järvi 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 phases in which they occur (Prior & Marcos-Cuevas, 2016). This way, managers can identify the pitfalls and adopt preventive or remedial action (Järvi et al., 2018).

The normative literature provides prescriptive guidelines and best practices for ensuring successful outcomes (Willumsen et al., 2019). From this perspective, value destruction may occur if the business strays from the guidelines and fails in the interaction process (Prior & Marcos-Cuevas, 2016). But Troilo, de Luca, and Guenzi (2017) argue that traditional strategic frameworks fail to 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 (Järvi et al., 2018). Of particular relevance for AI solutions are the lack of access to key data, inadequate information sharing (Vafeas, Hughes, & Hilton, 2016), and insufficient information-technology (IT) assets (Benaroch & Chernobai, 2017). The lack of suitable IT resources has been shown to "destroy value in a firm rather than simply fail to add any" (Arend, 2003, p. 280). Goldstein, Chernobai, and Benaroch (2011) go so far as to argue that a lack of functional IT resources can be even more harmful to a firm's value-creation efforts than data-protection failures.

In addition, value creation requires that firms embrace change (Järvi et al., 2018) and adapt their behavior accordingly (Homburg, Jozic, & Kuehnl, 2017). Digital technologies in particular require firms to change their behavioral models and how they interact with their stakeholders (Järvi 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 technologies.

5. Implementing the framework

We now discuss how the theoretical concepts previously presented may be used as a diagnostic tool (see Figure 2) to help managers tell when deploying AI solutions could result in value destruction for their firms. We illustrate the

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

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 (see Table 2). Starting with the input data, the business needs to have a channel to collect 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 control. When customers interact with the firm via Twitter, they may be unaware that they are interacting with a chatbot and that the data they provide will also be collected and analyzed by Twitter itself.

The chatbot also needs access to FAQ databases, inventory and other data, and the company's customer support team (Wilson & Daugherty, 2018). In addition, in order to personalize its answers, the bot needs access to a database of historical customer data, including past interactions and the customer's lifetime value or propensity to churn (Kietzmann et al., 2018).

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

Finally, the chatbot needs to perform a task. Four types of tasks are possible (Huang & Rust, 2018): mechanical tasks, such as delivering a scripted response based on keywords used by the customer; analytical tasks, such as reaching a conclusion about the type of problem faced by the customer; intuitive tasks, such as understanding why the customer is complaining; and empathetic tasks, such as trying to calm down an upset customer. Intuitive and empathetic tasks are harder than mechanical and analytic ones, even for very powerful AI solutions.

Chatbot tasks can be performed autonomously (e.g., providing delivery information) or through a member of staff (e.g., approving a refund), which requires connectivity. Chatbots that interact directly with customers require natural-language-generation abilities to produce replies intelligible to nonexperts and adapted to the circumstances of the complaint.

5.2. Predicting how value may be destroyed

Businesses have long adopted forms 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 complaints about delayed deliveries, returns and exchanges, and requests for compensation. In our scenario, the chatbot will be interacting with customers on Twitter, which means that the business can use an application programming interface (API) or

Table 2. Assessment of chatbot risks

Component	Connectivity	Cognitive ability	Imperceptibility
Input data: 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: Natural-language processing (NLP); Sentiment analysis; Result ranking	Needs access to contextual information to correctly assess sentiment	NLP offers flexibility but increases likelihood of error, and requires processing capability	May be unable to assess result ranking
Output decision: 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

processing application to automate collection and analysis of tweets, profile data, and metadata (e.g., location).

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 interact with these or other fraudulent accounts, 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 should result in comical or offensive replies, as with Microsoft's Tay chatbot.

Once the chatbot establishes that it is interacting with a person, it then 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 AI is now capable of collecting unstructured data, many businesses do not use such technology, whether because of limited budgets or incompatibility with legacy systems. Other problems can crop up, too. Some unstructured data may be unusable, such as images with too low resolution (Solon, 2018a); or the chatbot may be unable to draw on all available data sources owing to a lack of processing power (Agarwal, 2014). Moreover, while the chatbot may be programmed to detect common sentiment features indicating 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).

As for the algorithm, it is crucial that it use a technique that matches the type of problem at hand. 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 lack the necessary technical or nonlinear-thinking skills required (Merendino et al., 2018).

The algorithm will need constant updating (Khan et al., 2016) to reflect changes in regulations, new product lines, or recent promotional activities. Otherwise, it will lose the contextual relevance necessary to address customers' complaints. Moreover, in addition to its mathematical rules, the algorithm must operate according to certain assumptions about the world, and these will need to be updated often (Khan et al., 2016). If the programmers make 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.

Value can be destroyed in still other ways and circumstances. Customers who do not realize they are interacting with a bot could grow frustrated if, for instance, the chatbot 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 example, an FAQ database may require staff to record all questions, including unusual ones. If funds to support this effort are limited or if the employees do not record with sufficient diligence, the resulting database will be incomplete. Chatbots also need access to adequate historical data, and the database of possible solutions should be representative of the full range of customer problems.

If a 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). And unsupervised learning can create self-reinforcing feedback loops, quickly becoming so complex that even the people who created the algorithms can no longer explain how they work (Hudson, 2017).

5.3. Assessing what value could be destroyed

In recent years, companies have increasingly used 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). But customer complaints are also a critical point for customer satisfaction and recovery following a service failure (Istanbulluoglu et al., 2017).

Concerning the value of an algorithm's output, how should a good outcome be defined? On this question, the interests of the business and of 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, a chatbot is unlikely to be so nuanced. AI deals best with mechanical or analytical tasks, but struggles with intuitive or empathetic tasks (Huang & Rust, 2018). This is a problem in complaint management, where intuition and empathy are key to understanding the type of outcome sought by the client and to dissolving tension (Istanbulluoglu et al., 2017).

Firms may be tempted to fully automate their conversations with customers in order to save money, but this can yield unsatisfactory results. For instance, instead of solving the problem quickly, the chatbot may end up creating confusion or delaying 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 (Khan et al., 2016), though these too need to be adapted to the person using the visualizations and complemented with training.

6. Implications for managers

Having discussed how AI solutions can create problems for the businesses deploying them, we now reflect on the obstacles businesses are likely to face when solving those problems. First, given the plethora of issues vying for their attention, managers need to be able to determine which they should devote their attention to (Davenport & Beck, 2002). To do that within the context of AI, 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 effects of each detected problem can be ranked using a Likert scale, thus showing which events are most likely and most severe. Then, managers can produce a visualization of the source and effects of each risk to help demonstrate the potential for value destruction to others in the firm (Lowy & Hood, 2004). For instance, an untested ML model would represent a potential weakness. If the AI-powered solution in which that model is employed is not connected to other components and is used with few customers, it represents only a slight risk. But if the solution is deployed quickly, or if it provides advice in regulated industries such as financial services, small mistakes can easily grow into big problems.

Second, preventing or addressing the problems identified can be costly. AI and ML can represent a heavy initial investment and require continual maintenance. Even the simplest AI solution requires heavy initial investment in training (Solon, 2018a). It also needs processing power, access to various databases, and regular updating, all of which are costly. AI and ML also require specialist skills that most businesses lack. 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 difficult choices about value and values (Hudson, 2017). It requires that businesses confront ingrained operational biases that limit the quality of their input data, training data sets, and algorithms. Businesses need to decide what type of accuracy is most important and whether they prefer to incur false negatives or false positives; they also need to define fairness and decide whom they are most concerned with treating fairly. The outcomes produced through AI may be highly consequential for the firm and its customers. For example, predicting people's sexuality may sound innocuous when it comes to personalizing an advertisement, but this could result in one group's not being given the same opportunities as others, which is a form of discrimination. It could also increase some customers' social and economic vulnerability or even 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.

If businesses wish to benefit from using AI and ML tools, they will need a sophisticated understanding of the tools, a careful analysis of the risks, and sufficient initial investment in order to avoid inadvertent value destruction. The framework presented in this article 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). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, *25*(3), 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), e40014.

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, September 12). What every CEO needs to know to succeed with AI. *McKinsey*. Available at https://www.mckinsey.com/business-functions/mckinsey-digital/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. Available at https://doi.org/10.1155/2016/9139380

CCW. (2017). *Chatbots: Everything you need to know* [White Paper]. Available at https://www.customercontactweekdigital.com/events-customercontactweek/downloads/chatbots-everything-you-need-to-know

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). London, UK: McGraw-Hill.

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, J. F., Jr. (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 the 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, July 20). Technology is biased too. How do we fix it? *FiveThirtyEight*. Available at 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, March 2). Algorithms: The new means of production. *Digitalist*. Available at 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, March 25). Learning from Tay's introduction. *Microsoft*. Available at 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*(2) (XXX—XXX).

Lepak, D. P., Smith, K. G., & Taylor, M. S. (2007). Value creation and value capture: A multilevel perspective. *Academy of Management Review, 32*(1), 180—194.

Levin, S., & Wong, J. C. (2018, March 19). Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. *The Guardian*. Available at 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*. Available at https://engineering.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 problems and make better decisions*. London, UK: Jossey-Bass.

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

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

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). Playing Atari with deep reinforcement learning. *Cornell University*. Available at https://arxiv.org/abs/1312.5602

O'Hear, S. (2017, January 16). Fraugster, a startup that uses AI to detect payment fraud, raises $5M. *TechCrunch*. Available at 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*(2) (XXX—XXX).

Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York, NY: Free Press.

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* (3rd ed.). London, UK: Pearson.

Schwab, K. (2016). *The fourth industrial revolution*. Geneva, Switzerland: World Economic Forum.

Sharpe, A., & Raj, S. (2017). Using AI to determine queer sexuality is misconceived and dangerous. *The Conversation*. Available at 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., et al. (2017). Mastering the game of Go without human knowledge. *Nature, 550*, 354—359.

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

Solon, O. (2018 July 6). The rise of 'pseudo-AI': How tech firms quietly use humans to do bots' work. *The Guardian*. Available at https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies?CMP=Share_AndroidApp_Direct_Message

Solon, O. (2018, May 8). Google's robot assistant now makes eerily lifelike phone calls for you. *The Guardian*. Available at 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, June 10). AI cancer detectors. *The Guardian*. Available at 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*, 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. *Cornell University*. Available at https://arxiv.org/abs/1703.03107

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.