WP-1211-E
February 2019

Managing Artificial Intelligence within a Digital Density Framework[1]

JAVIER ZAMORA
Senior Lecturer of Information Systems, IESE

Abstract

AI is a new technology that enables organizations to leverage their data to create new value propositions. At the same time, the introduction of AI into the organization translates into new challenges, since the data—both as input and as output of AI models—require specific governance. This chapter outlines the holistic guidelines that general management should follow to achieve good AI governance within a digital density framework.

Keywords: Artificial Intelligence; Digital Density; Governance; Value Proposition; Organizational Model; Business Model

[1] This working paper is an earlier version of a chapter in the book "The Future of Management in an Artificial Intelligence World", J. Canals and F. Heukamp, eds., 2019, Palgrave Macmillawn, Houndmills, Basingtoke, Hampshire.

Last edited: 5/21/19

Contents

1. Introduction

The introduction of artificial intelligence (AI) into an organization should not be considered as a new technology in isolation but coupled together with other new technologies, such as social media, mobile, cloud computing, big data and Internet of things (IoT), among others. Together, they constitute mere manifestations of an environment with an exponentially increasing digital density (Zamora 2017), which I defined as the percentage of connected data that is available per unit of activity—a unit of activity being a country, a region, an industry, an organization or a business unit. In other words, digital density is an indicator of how many of the processes that are conducted in a given unit of activity are based on data that can be accessed remotely (i.e., connected data). In this sense, connected data become an abstraction of the physical entity itself, which can be remotely observed, monitored and/or controlled.

This increase of digital density is often used to gauge an organization's potential to generate new business models. As digital density intensifies, the once sharply defined lines between the digital and the physical worlds begin to fade, forging a new, blended environment, in a process known as digital transformation. Therefore, we should not consider AI as a mere technological infrastructure. AI has an impact on the business model (Casadesus-Masanell and Ricart 2011) by allowing new value propositions and, on the other hand, an impact on the organization in terms of governance, capabilities and cultural change. **Table 1** summarizes the business and organizational dimensions that a manager should take into account when introducing AI technology into an organization.

Table 1
Business Model and Organizational Model Dimensions of AI

Value propositions based on AI	Automation
	Anticipation
	Coordination
	Personalization
AI challenges	Privacy
	Integration
	Reliability
	Security
AI governance principles	Fairness
	Accountability
	Transparency
	Ethics
	Practical wisdom

Source: Prepared by the authors.

The scope and timing of AI's impact varies from industry to industry. For this reason, we will use examples of different sectors (e.g., healthcare, financial, retail public sector, etc.) to emphasize the different degrees of complexity and risk involved when using new value propositions based on AI. We will first review *why* AI is today a reality in those sectors, identifying the new sources of (big) data. Secondly, we will answer *what* kind of new value propositions based on AI are feasible in different sectors. We consider examples of AI in the context of four types of interactions; namely, automation, anticipation, coordination and/or personalization of interactions. Finally, we will address the new challenges in terms of privacy, integration, reliability and security that AI implementation (*how*) poses to any organization.

As connected data—as both input and output of AI algorithms—become one of the main assets of organizations, we need to understand the best way to incorporate this technology into the firm's business model. More often than not, deploying any new technology in an organization requires a transition period during which two modes coexist: a learning mode through pilots; and an earning mode, by executing the current business model. During this transition period, the organization should identify the newly required operational capabilities to successfully manage AI technology. In addition, general management (Andreu and Ricart 2014) should be aware of the new managerial challenges they will face as AI becomes more present in their organizations.

Firstly, organizations will face important issues regarding the fairness of AI models, depending on the bias introduced by the training data set. Secondly, as AI is going to be integrated in the decision-making process, issues about accountability will have to be faced in the event of undesired outcomes. Thirdly, general managers will trust those AI systems only if those systems are transparent to them instead of becoming merely a black box; that is to say, systems that explain themselves regarding how they reach certain recommendations. Fourthly, AI should take into account any decisions made on ethical issues based on the values (utility function) when those algorithms are designed. Last but not least, the use of AI must also be guided by the good judgment of the general management, who must act on the basis of what is right for all stakeholders, based on a practical wisdom that is aligned with the organization's mission.

This chapter begins by introducing AI in the context of the digital density framework, which includes three different dimensions: the technology model, the business model and the organizational model. Then, using examples of several industries, we illustrate the new kind of value propositions using AI that are feasible today. These new value propositions are the result of combining AI technology in one or more of four types of interactions: automation, anticipation, coordination and personalization. Next, we address the AI challenges in organizations in terms of the privacy, integration, reliability and security that these new value propositions based on AI pose to the organization. Following these challenges, we identify new capabilities needed to implement AI successfully in the firm. Thereafter, we identify the AI governance principles in terms of fairness, accountability, transparency, ethics and practical wisdom that a general manager should be aware of, and act on accordingly, regarding the externalities of AI beyond its impact on their business models. Finally, general management should manage AI in a holistic way in the organization, not only by leveraging the benefits of AI in the design of new value propositions but also understanding AI's current limitations, in order to address new challenges and minimize the negative externalities of the use of AI with customers, employees and society at large.

2. AI within the Digital Density Framework

Although the origins of AI (Zamora and Herrera 2018) as a new discipline date back to the year 1956, it has only recently gained momentum and many industries have started looking at AI as a promising technology. The AI renaissance as a viable technology happened mainly due to the confluence of three factors. The first factor is the increase in computational power and the decrease in its cost, as a direct consequence of Moore's Law. The second factor is the availability of huge data sets (big data) derived from a hyperconnected digitized world. And the third factor is the advance in the scalability and performance of AI algorithms.

In the context of the digitalization of organizations, people and even things, we should not consider AI as a technology in isolation, but together with other technologies, such as social media, cloud computing, mobile, big data, IoT and blockchain, as a manifestation of a world with an exponential increasing digital density (Zamora 2017). In other words, as many more processes of organizations, people and things get more and more connected, this translates into a growing digital density and begins blurring the frontiers between the physical and digital worlds. This new scenario, where the physical and digital worlds are indistinguishable, is the underlying driving force of the digital transformation that many organizations have been undergoing in recent years. Therefore, AI is also a technology that leverages this scenario of high digital density by turning the connected data into new sources of value creation and capture for the organizations.

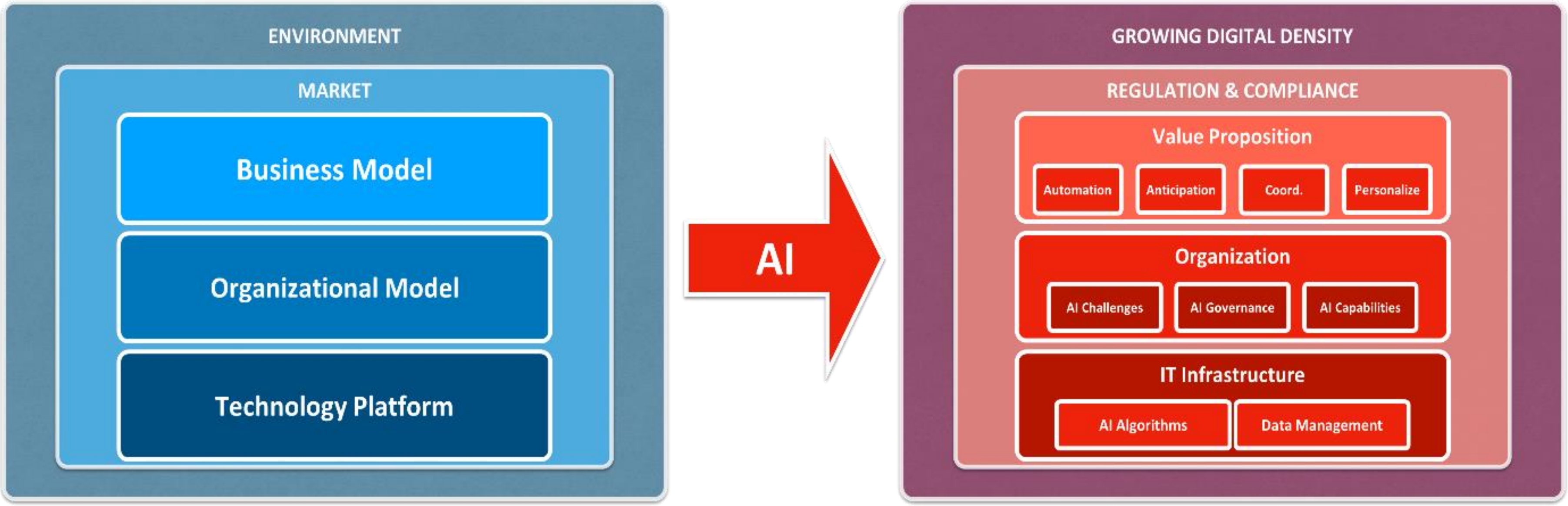
Andrew Ng, Adjunct Professor at Stanford University and a worldwide expert in AI, considers AI to be a general-purpose technology, in the same way as electricity has been. In other words, AI has the potential to redefine many industries in the same way that electricity redefined industries at the beginning of the 20th century or, more recently, the Internet changed the way many companies compete. However, in the same way that a company does not become an Internet company just by creating a web page, a company does not become an AI organization by the mere acquisition and introduction of AI systems in their internet technology (IT) portfolio. To that extent, AI, as well as other new technologies involved in the digital transformation process, should be considered in a holistic way when considering its impact on different dimensions (see **Figure 1**):

- Technology platform
- Business model
- Organizational model

In the specific context of AI, the technology platform refers to the required IT infrastructure, which mainly comprises of a collection of AI algorithms (Zamora and Herrera 2018) that today—more often than not—are machine-learning algorithms performing mainly prediction and/or classification functions. However, the competitive advantage does not reside in owning those algorithms, since the majority of them are available to many organizations, but in having the data to train and test the algorithm to build and validate a model to be used later with the new data. Consequently, in a world of high digital density, the data (Zamora, Tatarinov and Sieber 2018) become one of the fundamental assets of the organization. Therefore, the IT infrastructure also comprises all the information systems necessary to have efficient data management (i.e., capture, curation, search, protection, etc.).

Figure 1

Framework for Digital Transformation (Left-Side) and Its Application in AI (Right-Side)



Source: Prepared by the authors.

The next dimension is the business model, which refers to the underlying logic and dynamics (Ricart 2012) of a business to create and capture value. One integral component of the business model is the value proposition (Osterwalder and Pigneur 2010), or the products and services that create value for a given customer. In this regard, new value propositions are enabled by using AI technology in four types of interaction:

- Automation, or using AI to automate existing processes by removing manual steps to achieve cost reductions.
- Anticipation, or using AI for prediction or recommendation purposes.
- Coordination, or using AI to coordinate in an intelligent way a multitude of actors who participate in the creation of the value proposition.
- Personalization, or using AI to customize the value proposition for a given customer.

AI processes data to build the new value proposition (Zamora 2017) using a combination of some of these four types of interaction.

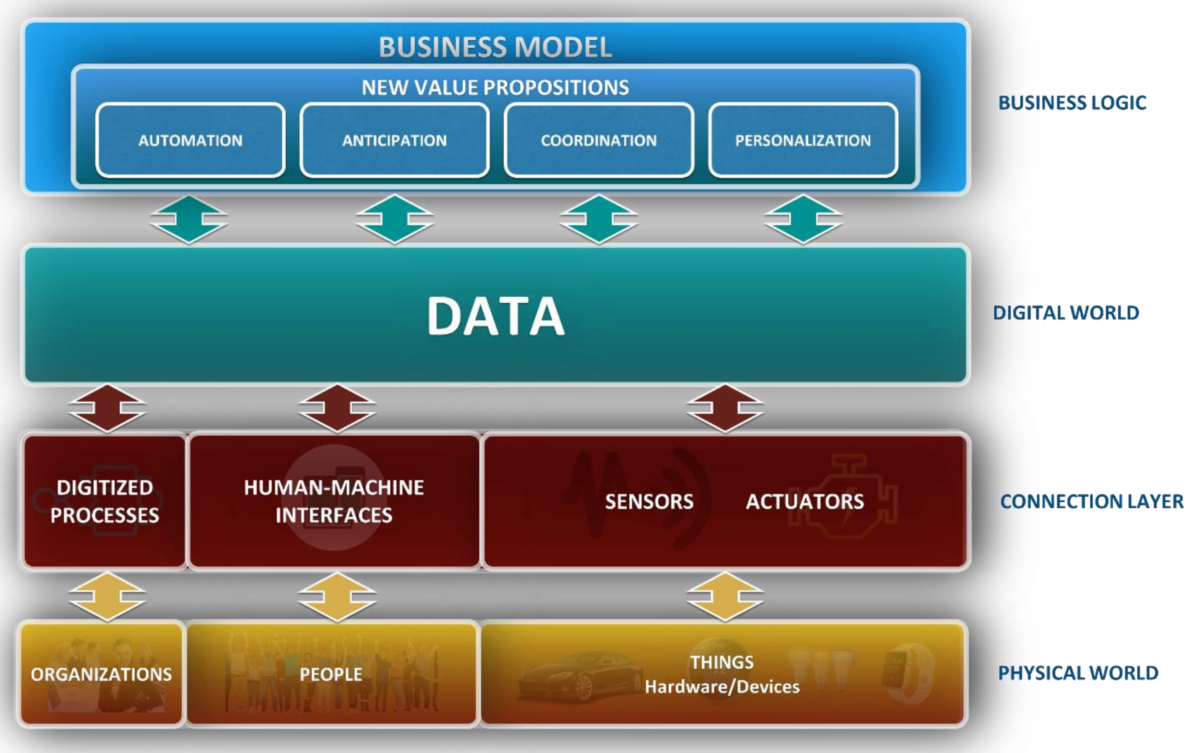
The organizational model dimension refers to how AI has an impact inside the organization (Káganer, Zamora and Sieber 2013). This includes several aspects. On the one hand, how organizations start using AI in pilots for a learning purpose to explore the potential of the technology, as well as later scaling successful pilots into production as an integral component of the execution of a given business model. On the other hand, AI calls for new capabilities, both at the operational level (e.g., data scientists) and managerial level, to address the new challenges regarding privacy, integration, reliability and security. Moreover, additional managerial capabilities are required for the governance (Andreu and Ricart 2014) of AI inside an organization, given a regulatory and compliance framework where the company develops its activity. Specifically, this governance should address issues related to the fairness, accountability, transparency, ethics and practical wisdom when an organization offers a new value proposition based on AI.

This chapter will focus on the business model and organizational model dimensions.

3. New Value Propositions Using AI

In **Figure 2** we show the digital density architecture that interconnects the physical world with the digital world, together with the business logic. The bottom layer represents the physical world, consisting of organizations, people and things. Above the physical layer we have the connection layer, which relates the physical world to the digital world. Organizations have traditionally been connected to the digital world by digitizing their processes (e.g., enterprise resource planning (ERP), customer relationship management (CRM), etc.), and people are connected to the digital world through human machine interfaces (e.g., web, app, voice, etc.), while things are being connected either through sensors to read their state (e.g., position, temperature, speed, etc.) or through actuators to change their state (e.g., turn it on, accelerate, etc.). On top of this connection layer, lie the connected data that represent the physical world (Zamora 2017).

Figure 2
Digital Density Architecture



Source: Prepared by the authors.

As digital density increases, the data layer better represents the physical world. In this scenario data are becoming the main asset of the organization, since data have become the raw material for creating new value propositions and, in turn, for building new business models. Metaphorically speaking, if data are considered the new oil of the economy, then AI becomes one of the engines that transforms these data into new value propositions.

One example of increasing digital density in the health sector is the Biobank Data in the United Kingdom, where data have been collected from over 500,000 people for the last 30 years, including their medical history, imaging, genetic data—via the European Genome-phenome Archive (EGA)—and physical and medical activity through mobile monitoring. Health researchers, using AI technology, are working with this data repository to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses. Their work is being translated into the deployment of intelligent systems in healthcare, where doctors can now map a patient's data, including what they eat, how much they exercise and what is in their genetics; cross-reference that material against a large body of research to make a diagnosis; access the latest research on pharmaceuticals and other treatments; consult machine-learning algorithms that assess alternative courses of action; and create treatment recommendations personalized to the patient.

As we mentioned above, the connected data derived from a high digital density environment can be used to build new value propositions as a combination of one or more of four types of interaction: automation, anticipation, coordination and/or personalization. Although the majority of the new value propositions are the result of the combination of more than one type of interaction, the following examples are categorized under the type of interaction, which is more prevalent in the value proposition.

3.1. Automation Using AI

Traditionally, organizations have been connected to the digital world by digitizing their different processes. The majority of these processes could be automated, since they could be described by workflows handling limited and well-defined cases that could be implemented in enterprise software applications (e.g., ERP, CRM, etc.). The drastic reduction of computational cost due to Moore's Law has been behind the popularization of this type of software, enabling massive digitalization of companies.

However, some organizations' activities require a more sophisticated automation, because they imply an almost infinite number of scenarios (exceptions) that traditional software cannot deal with. In those cases, robot process automation (RPA) or AI workers can be used, where AI systems watch/observe the activity of a worker and learn from her/his actions, taking advantage of AI systems' large capacity to remember. For instance, AI can be very efficient in the legal world, where traditionally lawyers spent hours searching through documents and looking for evidence (i.e., responsive documents) for a given trial. AI can automate most of the process by pre-classifying the documents into two categories, separating the ones that are not responsive from the ones that might be responsive and should be submitted to a lawyer for final classification. Similar application of AI can be found during mergers and acquisitions (M&A) procedures when looking for clauses in all contracts that can imply future liabilities. For instance, clauses in clients' contracts that can be terminated in the event of an acquisition.

The previous examples are doable because of the advances in the natural language processing (NLP) techniques for classifying purposes. NLP can also be used to enhance productivity by introducing virtual assistants to automate interactions with customers, which is being increasingly used with chatbots in the finance and retail industries. The automation process based on implementing AI in the health sector is represented by algorithms managing data related to medical records, analyzing medical tests, X-rays and CT scans. Sense.ly, a health industry start-up, developed a digital nurse called Molly that helps patients with symptoms that do not require a visit to a physician. Boston Children's Hospital similarly uses Amazon's virtual assistant Alexa to advise parents about children's treatments or whether symptoms require a visit to a doctor.

Using AI (and other digital technologies) for automation purposes brings cost down, since less manual work is needed. Frequently, automation is the first step that many organizations take in the context of high digital density as it directly replaces previous manual processes by digitalizing them. However, as technologies spread to more organizations, businesses should focus on other interactions beyond automation (e.g., anticipation, coordination and/or personalization) to maintain a competitive advantage.

3.2. Anticipation Using AI

As digital density increases exponentially, organizations can take advantage of the generated big data by anticipating patterns and trends that the data reveal. Therefore, anticipation in this context means the ability to make predictions; that is to say, using existing data to generate new data that organizations do not have. In the same way that automation was widely used by companies as the result of having affordable computation, the inclusion nowadays of anticipation in many of the new value propositions is the consequence of prediction becoming cheap (Agrawal, Gans and Goldfarb 2018).

We can find an example of a value proposition using anticipation in the aeronautics company Rolls-Royce, which can inform an airline when a plane that is landing needs preventive maintenance in advance of the scheduled date. In this way the airline can avoid unscheduled stops, leading to substantial savings—given that an unscheduled airplane stop due to technical problems costs approximately US$10,000 per hour. Rolls-Royce receives real-time operating data, from more than 25 sensors per plane, for each of the more than 12,000 Rolls-Royce engines operating worldwide. By cross-referencing with records of problems with other engines and applying a predictive algorithm, the company can predict technical problems in specific engines even before they appear. Rolls-Royce's use of AI implied migrating its business model: instead of selling a product (an engine), it offers its customers, the airlines, a service based on the number of airplane engine hours without unscheduled stops (Zamora 2016).

Another interesting example of anticipation in the health sector is Cardiogram, an American company that offers a mobile app that acts as a personal healthcare assistant. Cardiogram leverages the data coming from personal wearable devices like the Apple Watch or Android Wear, not only to track sleep and fitness activity but also to detect atrial fibrillation. Atrial fibrillation is a type of heart arrhythmia that causes more life-threatening strokes than any other chronic heart condition, and in many cases is undiagnosed, since continuous heart monitoring is needed. In 2016, Cardiogram collaborated in the mRhythm Study (Health eHeart 2018), together with the University of California, San Francisco (UCSF), training Cardiogram's deep learning algorithm, Deep Heart, with 139 million heart measurements coming from 9,750 users. The result of the study showed an accuracy of atrial fibrillation detection higher than Food and Drug Administration (FDA)-cleared wearable electrocardiogram (ECG) devices. Cardiogram is currently deploying Deep Heart outside of the mRhythm Study to offer it to all the Cardiogram app users.

Notwithstanding, a paradigm of using anticipation in new value propositions is IBM's Watson, which has gained a lot of notoriety over the last few years. Watson has AI capabilities to process natural language and generate hypotheses and learning based on evidence. Initially, IBM created the Watson project in 2006 to find out whether a supercomputer could compete with humans' decision-making capabilities when faced with complex challenges. In 2011, after five years of work and training, Watson was sent to compete on *Jeopardy!*, a television game show, against two of the best contestants from previous rounds. After several rounds of play, Watson won out.

In the wake of the media reporting, in 2013, IBM set up a partnership with the Memorial Sloan Kettering Cancer Center in New York to use Watson for treatment decisions in cases of lung and breast cancer. When making decisions, Watson processes more than two million pages of research articles from medical journals, in addition to an archive of more than 1.5 million patients' medical records. To aid physicians in the treatment of their patients, once a physician has posed a query to the system describing symptoms and other related factors, Watson first parses the input to identify the most important pieces of information; then mines patient data to find facts relevant to the patient's medical and hereditary history; then examines available data sources to form and test hypotheses; and finally provides a list of individualized, confidence-scored recommendations. Based on all of this information from tests performed in previous cases, the percentage of successful treatments prescribed by Watson for lung cancer cases is 90%, much higher than the 50% achieved by doctors (Steadman 2013).

Watson's advantage over human beings is obviously its ability to absorb information. In fact, in a 2017 study (Wrzeszczynski et al. 2017) IBM Watson took just 10 minutes to analyze a brain-cancer patient's genome and suggest a treatment plan in comparison with the 160 hours that human experts needed to make a comparable plan. Compared with traditional software, which is based on predetermined algorithms and always yields the same output if given the same input, Watson uses technology based on machine learning, where the algorithm is adapted as a result of the learning that occurs during the training process. In Watson's case, this consists of looking at a treatment's effectiveness among the patients who have received it.

In the financial sector, AI can be applied to improve the credit scoring of a given client. Traditional algorithms do not do well at predicting the deterioration of a credit score over the years. In a 2010 study (Khandani, Kimz and Lo 2010), an alternative to the traditional credit score was used as a risk classifier using machine learning. It used a 1 terabyte dataset consisting of everyday transactions (e.g., via credit cards, ATMs, etc.), credit bureau data, and the account balance for a subset of a commercial bank's customers, which accounted for 1% of the data generated by the bank for the period between January 2005 and April 2009. The study showed that machine learning performed better as a risk classifier than traditional credit score algorithms.

Some good clients could previously have been rejected because of a low score using the traditional algorithms and, conversely, some bad clients could have been accepted since they were getting a high score. By using this alternative method, banks can get both a cost reduction, by better identifying riskier operations, and, at the same time, generate new business with people who otherwise did not have a chance to become their clients.

In all the examples above, the use of AI in the form of an anticipation interaction allows organizations to predict the state of the physical world, which is a critical factor in developing new value propositions. The main business drivers of anticipation are: description, prediction and prescription. First, we can use AI to describe a complex process that otherwise would not be evident to detect—as though using a digital microscope—as in the case of Cardiogram's app for detecting atrial fibrillation. Second, we can use AI to predict future patterns based on current conditions, as in the case of the Rolls-Royce engines or the risk classification for consumer credits using machine learning. Finally, we can use AI to prescribe or recommend a course of action; for example, IBM Watson recommending a specific oncologic treatment for a given patient.

3.3. Coordination Using AI

Traditionally, organizations have offered their value propositions operating inside the boundaries of the linear value chain of a given sector (e.g., automotive, banking, etc.). This situation derived from the high cost of transactions (e.g., coordination, production, etc.) that made it unfeasible to do it otherwise. As a result the products and services were fully made and controlled by the organizations themselves, with the participation of the providers present in their value chain. However, as digital density increases, it becomes possible to redefine how customers' needs are met beyond what is provided by the traditional value chain. In other words, new value propositions can now be the result of the coordination of disparate and multiple actors (i.e., organizations, people and things).

As the number of actors involved in a given value proposition increases, the complexity of the coordination increases substantially, since the number of possible interactions and learning opportunities grow in an exponential combinatorial fashion. In these scenarios, the use of distributed AI (DAI) can be very helpful to aid building new value propositions. DAI systems consist of autonomous, intelligent actors, called agents, that are physically distributed and often on a very large scale. One example could be the application of DAI in calculating the optimal routing of a large fleet of vehicles in a mobility platform (e.g., Uber, Cabify, etc.).

In all of these examples the use of AI for coordination interactions allows different actors to work together in new value propositions, without the constraint of belonging to the same traditional linear value chains and without the limitations of size (i.e., number of actors) or physical location. Nowadays, the use of AI for coordination interactions is not as prevalent as in the case of the other three types of interactions (i.e., automation, anticipation and personalization). Nevertheless, as we move to a hyperconnected world of high digital density, organizations, people and things could be coordinated regardless of their number and physical locations. For instance, the Chinese city of Guangzhou—with a population of 16 million people, more than 16,000 km of roads and a daily flow of 3.5 million vehicles—uses the program City Brain, based on AutoNavi, Alibaba's traffic management system. City Brain (Alibaba Cloud 2018) allows the Guangzhou Traffic Police Smart Center to analyze big data coming from video feeds, social media and traffic information to optimize traffic signals and reorganize their road network in real time.

Therefore, as digital density continues to increase, it will create a new scenario of collective intelligence, where people and computers (i.e., computation associated with connected things and/or organizations) might act collectively more intelligently than any person, group or computer (Malone 2018), enabling another level of value propositions based on AI.

3.4. Personalization Using AI

Until recently, organizations were competing in their markets either using pricing or differentiation strategies (Porter 1979); that is to say, by competing on price in a mass market or by developing products for a specific niche market. When digital density increases companies can create a fully personalized offering for a high volume of different customers (Anderson 2006), based on the data reflecting the habits and preferences of an individual consumer. AI is used in the interaction of personalization to predict the right value proposition for a given customer, based on the collected data reflecting her/his habits and preferences.

One example of applying AI in a personalization interaction is the US auto insurance company Progressive, where customers were able to opt-in to plug a small device called Snapshot into the car's onboard diagnostics (OBD) or by installing the Snapshot app on their smartphones. Snapshot tracks their driving behavior (e.g., how they turn the wheel, how they brake, etc.) and sends the data back to Progressive. The Snapshot program had collected more than 13 billion miles of driving data by 2016. All these data are processed by Progressive's partner, H2O.ai, using predictive analytics. In this way, the company becomes more efficient in its operations (e.g., managing claims, detecting fraud, improving analytics, etc.) and, at the same time, personalizes the customer experience. Those customers who voluntarily decide to share the data collected by Snapshot for the period of the first insurance policy, which is normally half a year, get a personalized insurance rate based on their actual driving rather than the standard car insurance criteria—such as age, car model or area of residence.

Another application of AI as personalization interaction can be found in precision medicine, by tailoring a medical treatment to the individual characteristics of each patient. Precision medicine involves analyzing the way that patients' various biological characteristics interact with multiple pharmaceutical molecules in order to better match drugs to improve patients' heath. However, precision medicine has not become a reality due to the unaffordable costs associated with the required combinatorial explosion of clinical trials that would be needed. For this reason, pharmaceutical companies have traditionally been offering standard care for a hypothetical, average patient. The US company GNS Healthcare processes millions of data points of all types–electronic medical records, genetic, proteomic, genomic, claims, consumer, laboratory, prescription, mobile health, sociodemographic, etc.–to model patient response to treatment *in silico*; that is to say, applying computer simulation instead of clinical trials. GNS Healthcare, using machine learning, reconstructs complex diseases into computer models, which allows pharma companies to simulate real-world scenarios, increasing the speed of discovery of new drugs from years to months.

In the previous two examples of Progressive and GNS Healthcare, the use of AI in personalization interactions allows organizations to create a specific and affordable value proposition based on customers' needs. In many of those cases, the personalization interaction is also related to the anticipation interaction. For instance, the traffic app Waze uses an anticipation interaction to leverage the big data received from the drivers for prediction purposes, whereas the personalization interaction focuses on the personal data (i.e., location of a given car) as the input to change the behavior (i.e., a driver changing his route following Waze's recommendations). Another example is Amazon's recommendation engine analyzing all the client transactions and clustering them in groups of clients with similar behaviors and tastes (i.e., anticipation) to recommend potential products to a given customer (i.e., personalization). As AI algorithms for anticipation and personalization continue improving, Amazon can eventually switch its business model from buying and then shipping to shipping and then buying, as was hinted at by the patent Amazon filed in 2013 for what it termed "anticipatory shipping."

4. AI Implications on Organizations

In the previous section, we saw that AI represents an important technology in terms of building new value propositions as a combination of automation, anticipation, coordination and personalization interactions. However, the introduction of AI in a company's business model also has a big impact on its organizational model. In this section, the AI implications are analyzed at three levels: firstly, identifying the specific challenges that arise when an organization implements new value propositions based on AI; secondly, focusing on new capabilities that an organization needs in order to integrate AI technology; and thirdly, stressing the importance of some AI governance principles when data become the critical asset for any business.

4.1. Challenges when Introducing AI

Beyond any doubt an increase in digital density enables organizations to leverage a lot of benefits by creating new value propositions, as described in the previous section. However, this scenario also creates new challenges, specifically in the context of AI, where organizations need to recognize and address certain issues that may arise in relation to:

- Privacy, related to the amount of personal data required to train the AI algorithms
- Integration, related to the ownership and use of data used for value propositions that is obtained as a result of coordination between multiple parties
- Reliability, related to the quality of AI model outcomes
- Security, related to the vulnerability of AI models against cyberattacks.

4.1.1. Privacy

However, better personalization implies collecting and storing more and more individual data from the customer, and the challenge of privacy is therefore incremented. A near-future scenario may involve a company that is able to predict, through biosensors, the likelihood of a person to develop a serious illness. Hence, a health insurance provider could potentially discriminate against a customer by denying coverage for those with higher risks. This personalization-privacy tension could even be found in a simple vacuum-cleaner robot that creates a map of an apartment to know which parts have been already cleaned. For this reason, consumers are only going to embrace these new value propositions if they trust the organization providing them and it is transparent in its use of personal data. First steps have been taken towards increased levels of personal data privacy by introducing new regulation and compliance requirements, such as the General Data Protection Regulation (GDPR) adopted by the European Union (EU), which also limits what kind of data an organization can use to train its AI algorithms.[2]

4.1.2. Integration

In a high digital density world, value propositions are often the result of the cooperation of multiple organizations within richer ecosystems. For this reason, such partnerships require establishing data clauses regarding the ownership and limitation of data usage by participating companies. In the previous example about predictive maintenance for aircraft engines, Rolls-Royce sees its operational value in the aggregated data from all the engines, while the airline and/or the aircraft manufacturer, only have access to a fraction of the raw data collected and do not have access to the bigger picture. Industrial companies may not want to share their data for compliance reasons (e.g., medical machines) or competitive concerns, thus limiting the machine learning capabilities (i.e., reducing the training data set of the equipment operated by the organization). In order to reconcile industrial companies' concerns about sharing their data because of the concern of sharing valuable insights about their products, some organizations have decided to share the data for the sole purpose of using it as a training data set for AI algorithms. However, they retain ownership of the data in order to be protected from compliance and security risks.

4.1.3. Reliability

Reliability in the context of AI encompasses two aspects. On the one hand, the quality of the data that are used by AI algorithms is important and, on the other hand, the reliability of the software of those algorithms represents challenges. Ensuring data quality relates to data governance, which will be addressed later on in this chapter. Regarding the reliability of the AI algorithms, it is important to take into account not only their possibilities but also their limitations. Addressing data quality and algorithm reliability issues within businesses will help organizations discard investments without any return. It will also help them not to oversell the results of projects involving AI, in order to avoid disappointment and skepticism regarding AI's utility.

The AI algorithm learning techniques (Zamora and Herrera 2018) are based on training a model with data. Without these data, training is not possible and, therefore, it is not possible to generate any model. Many times, companies request evaluation about what quantity of data is needed to train a model. This depends on the particular case and the complexity of the algorithm. However,

[2] GDPR focuses specifically on protecting data and ensuring its privacy. Any organization operating within the EU will be obliged to gather data legally and under strict conditions and protect it from misuse by third parties, otherwise it will be fined. Organizations are required to use the highest privacy settings, so that the data does not become public. GDPR empowers individuals to challenge organizations to reveal or delete their personal data.

the right question to be addressed by business executives should not be related to the amount of data needed, but rather what problem can be solved with the data available. If an organization does not have a lot of data, it is advisable to use those algorithms that are more resistant to learning with less data but which, however, imply lower confidence levels. For example, if a prediction model for weekly product sales in a retail store is needed, it is logical to request several years of sales data for each product per week, as well as the time series with potentially explanatory variables (e.g., holidays, weather, macroeconomic indicators, sales channel etc.).

An AI system with a high quality of prediction may stop performing as well at any time, or, to put it another way: an AI system properly trained with past data, can fail to be correct with the present circumstances if these circumstances change the reasons why the past cases occurred. For example, a customer who buys seasonal clothing for years in a store can change his or her habits with the appearance of a competitor that better meets his or her needs and at a better price.

Due to this limitation, it is essential in the current situation to create AI systems with inbuilt continuous learning but also with the ability to distinguish noise in the data; that is to say, events not relevant for the training but that can distort the results.

The degree of confidence of the resulting model built with a machine-learning algorithm can be measured, comparing the prediction of the model with what has happened in the past. The more cases that prediction coincides with reality, the greater the degree of confidence in the model will be. But in no case does the degree of confidence of a model fully guarantee its success in a future practical application, since many models suffer what is known as over-training; that is, the learning process has been too adapted to the data of training and, despite its high degree of confidence, the emergence of circumstances that generate relationships between data that differs from the past does not make the model work well in practice.

4.1.4. Security

One of the consequences of a hyperconnected world is that the surface of attack of the organization increases exponentially. For this reason, protecting the integrity of the data from cyberattacks is especially critical when data power the AI algorithms. Therefore, security in the context of value propositions based on AI goes beyond the type of cyberattacks (Sieber and Zamora 2009) we have experienced until now, something that is called adversarial machine learning. This kind of attack exploits the limitations of the neural network design used in many AI algorithms, which do not operate in the same way as a human brain does. A hacker (i.e., a malicious adversary) can manipulate the input data of an AI algorithm, either in the training or operating phase, to fool systems into seeing or listening to something that compromises the whole security system.

A training-time attack may happen at a stage of building a machine-learning model by using malicious data. An inference-time attack uses specifically synthesized inputs that affect the model. Some examples of AI being hacked may look innocent, like a neural network confusing a turtle with a rifle on a picture. However, some AI mistakes can cause greater disturbances, like a self-driving car not stopping at a stop sign, because it was partially covered by carefully crafted black-and-white stickers (Eykholt et al. 2018).

As more and more value propositions depend on using AI technology, it is important to identify potential risks related to adversarial attacks beforehand (i.e., secure-by-design principles) and build in certain defenses to protect them.

4.2. New Capabilities when Adopting AI

When an organization considers implementing new AI technologies, it is advisable to do it progressively in order to assess its suitability and prepare the organizational structure and capabilities both at the operational and managerial level to successfully integrate AI technology into its business model.

Quite often, organizations start with a pilot where AI can be used to create a new value proposition for automation, anticipation, coordination and/or personalization interactions. In these initial phases the emphasis should be put on iterative experimentation, building a minimum viable product (MVP) and tracking metrics, which allow testing initial hypotheses or success criteria. Once the new value proposition has been validated in the pilot project, organizations can start implementing it on a larger scale by replacing or improving existing processes. This will translate into cost reduction and/or generation of new sources of revenue. Moreover, the new data generated by the customer using these new products and services can be used in turn to discover or improve new value propositions—through the use of AI as a driver of innovation—closing a virtuous circle.

As a matter of fact, most AI algorithms are widely available. However, access to quality data is the main entry barrier for achieving a sustainable and competitive business model. For this reason, it is critical to define a strategy of continuous data acquisition (e.g., having a unified data warehouse), not only for AI training purposes but also as a source for future innovations. This requires the development of new operational capabilities in the organization related to the application and correct usage of AI techniques.

To that extent, many organizations incorporate new professional profiles, such as data scientist (Zamora and Herrera 2018). An ideal data scientist should have training in applied mathematics and a good knowledge of programming languages and database management. Moreover, a data scientist must be oriented towards practical results but with a great creativity component, especially when it comes to defining the data training for the AI algorithm.

However, organizations face some problems filling this position. Among the difficulties that organizations encounter when searching for data scientists are: domain knowledge of the business, communication skills, and understanding and identifying the different and heterogeneous repositories of company data. In order to overcome these limitations, a data scientist's role often serves in conjunction with other roles, such as business translators who interpret business challenges, points of improvement and opportunities, and translate them into proposals that can be implemented using AI.

In addition, organizations should also incorporate profiles specialized in searching for the necessary training data in a more efficient way. In some instances, availability of a unified data warehouse or any other centralized repository of information simplifies the task, but there is always the subsequent enormous search task among the different business attributes of which the training data set is composed. The knowledge that this profile must act in compliance with the regulatory requirements when processing personal data (e.g., GDPR) deserves a special mention.

As AI technology becomes a core technology in organizations' business models, new capabilities (Daugherty and Wilson 2018) related to the AI governance will be needed. These will be described in more detail in the next section.

4.3. Some Governance Principles with AI

As described in the previous sections, the introduction of AI technology as a part of new value propositions will face specific challenges of privacy, integration, reliability and security, as well as difficulties in acquiring the new capabilities necessary to manage them. However, since AI technology is used either to substitute or augment human activity, general managers should be aware of the impact of AI on their own decision-making process, on employees, on customers using their products, and on society as a whole. For these reasons, general managers should be aware and act accordingly regarding the externalities of AI, especially the negative ones, beyond its impact on their business models.

First, organizations will face important issues regarding the fairness of implemented algorithms depending on the bias[3] embedded in the training data set. Second, as AI algorithms are going to be integrated in the decision-making process, issues about accountability will have to be faced in the event of undesired outcomes. Third, general managers will trust AI systems only if these systems can be transparent and explainable (e.g., able to explain how certain outcomes were reached) instead of being merely a black box. Fourth, general managers should be conscious of the ethical implications of using systems based on the values (utility function) that the AI algorithms were designed with. Lastly, general managers should follow practical wisdom, acting on the basis of what is right for all stakeholders.

With the current state-of-the-art AI technology, the present-day problems that organizations are facing do not relate to the possibility that AI systems will achieve a super intelligence power (Bostrom 2014) overriding any human control but, on the contrary, are based on the shortcomings of today's AI technology, which lacks inherent human abilities such as generalization, abstract thinking and the ability to make analogies. In the words of Pedro Domingos, professor at the University of Washington and a leading AI researcher, "people worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world" (Domingos 2015). This means that AI algorithms do not yet have the capability of understanding things the way humans do, which may result in dangerous outcomes.

These AI limitations are directly related to the quality of the data used and the design itself of the algorithms. Therefore, organizations should know and trust their data before starting to use AI technology. "Data for Good" (Wing 2018) has been advocated for in the scientific and technology field as a guiding principle of data usage through the entire data life cycle. In the management world, a similar principle should be used when adopting AI in the decision-making process and integrating AI in the company's offering. This guiding principle of data usage and design algorithm design is also known by the acronym FATE (fairness, accountability, transparency and ethics) (FAT/ML 2018) (Microsoft 2018). In the specific context of general management, we propose to add an additional guiding principle, based on practical wisdom, to address the issues of:

- Fairness, related to the bias introduced by the training data set of the AI algorithms
- Accountability, related to the responsibility of the decisions based on AI models
- Transparency, related to the explainability of the AI models
- Ethics, related to the values on which the AI systems are built
- Practical wisdom, related to the good judgment of the general management as to whether to use or not use AI on the basis of what is right for all stakeholders.

[3] In this context, bias does not refer to its statistical meaning but to the inclination or prejudice for or against one person or group, especially in a way considered to be unfair.

4.3.1. Fairness

As we have seen in this chapter AI is a powerful technology, able to serve customers better and get more enhanced insights into the organization. However, these advantages should be realized while avoiding exposing people to any kind of unfairness as a result of potential biases introduced by the adoption of AI. In this context, fairness means that the adopted AI models must produce unbiased classifications or predictions. One trivial example can be found in the image search engine Google Images. For the query “CEO” it overwhelmingly produces images of men in suits, reflecting how, historically, women have had difficulties in accessing high executive positions with conditions equal to those of men. In this case, the training data set contains the gender bias of having many more men than women in CEO positions.

Since an AI model is the result of training an algorithm with data, the bias comes from either the data used or from the algorithm itself. However, algorithms reflect the values of the people who designed and coded them. This is something that we will cover below, considering the ethical implications of AI. For this reason, organizations should pay special attention to the bias contained in the training data set used by AI algorithms. Although many AI systems were made with the best intentions, these AI systems increasingly have a direct impact on people’s lives (O’Neil 2016). As a result, biased data should be a concern for many organizations, due to important liability consequences in areas like access to housing, law enforcement and employment.

A recent study (Buolamwini and Gebru 2018) evaluated the bias in three commercial facial recognition software applications (Microsoft, Face++ and IBM). The study showed that the software was right 99% of the time for photos of white men. However, the error rate was nearly 35% for images of women with darker skin. The study concluded that the gender and race biases were due to the data set used to train those software programs. As a matter of fact it was estimated that, in the training data set used by the facial recognition software, about 75% of the pictures were portraying men and more than 80% of them were white men. These results indicate that there is an urgent need to fix gender- and race-biased data if companies want to incorporate this type of software in their commercial offerings.

Notwithstanding, the AI model bias can have even more serious consequences, causing social stigmatization by stereotype reinforcement. This seems to be the case of the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) software, developed by the US company Northpointe, which is used across the US courts to predict recidivism of criminals. COMPAS has become a tool used by judges to guide them in producing sentences by identifying potential re-offenders in a future crime. According to a recent study published in *ProPublica* (Angwin 2016), COMPAS predicted the recidivism correctly 61% of the time. However, it showed a strong bias towards black people, who were labeled with a high risk of recidivism 44% of the time, but then finally did not re-offend. In contrast, white people who finally did not re-offend were labeled high risk only 23% percent of the time.

For those reasons, the quality of data utilized by algorithms should be addressed by managers in advance of AI development and deployment. Intel, in their Public Policy Opportunity white paper (Intel 2017), suggests mitigating bias by using verified algorithms, data models and well-curated training sets, performing extensive validation of AI systems and being alert to possible fairness implications from AI-based decisions. To that extent, as companies are using multiple sources of data to feed their AI algorithms, a data hygienist role (Daugherty and Wilson 2018) will be required. Her/his role will be to free the data of any noise or hidden bias.

4.3.2. Accountability

Accountability has always been a crucial concept in management. According to Peter Drucker (Drucker 1973), leaders in any organization are responsible and accountable for the performance of their institutions and also responsible for the community as a whole. Therefore, accountability implies the acknowledgement and assumptions of responsibility for actions, products and decisions within the scope of the management role. However, in the context of AI, accountability can be more challenging, since behind a decision or a product are also data and algorithms. One example is the 2015 study (Datta, Tschantz and Datta 2015) of Google Ads, using a tool called AdFisher, developed by Carnegie Mellon University, which runs experiments using simulated user profiles. The study revealed that ads for high-paying jobs were shown more to men than women on Google. Nevertheless, in this situation it is not clear who is accountable for this discrimination outcome. It could be the advertisers targeting the ad, or the design of the Google advertising platform, or the fact that men click more on this type of ad —which leads to an increase in the frequency of appearance of these ads to men, or even that there is more competition for advertising space to women—which results in a reduction in the frequency of appearance of these ads to women.

Previously in this chapter, it was mentioned how IBM Watson helps oncologists in the diagnosis and type of treatment of their patients. However, if Watson provides a treatment recommendation with fatal consequences for the patient, who will be accountable for this undesired outcome: the oncologist, the hospital, or IBM? This scenario could potentially happen, as pointed out by the health-oriented news website STAT (Ross and Swetlitz 2018). In July 2018 STAT reported “multiple examples of unsafe and incorrect treatment recommendations” in cancer treatments produced by Watson.

Reflecting on these issues of accountability, the Future of Life Institute held the Asilomar (Future of Life Institute 2017) Conference on Beneficial AI in 2017, creating a set of 23 guidelines for AI researchers. These guidelines can be extended to managers of organizations, as they are the main stakeholders in the moral implications of the use, misuse and actions related to AI technology. Organizations should commit to a responsibility to shape those implications and adopt internal policies consistent with external social criteria. Companies should implement the necessary systems and training programs for managers to use AI tools and be held accountable for the outcomes.

4.3.3. Transparency

The third guiding principle in the use of AI is transparency. Organizations will only implement AI systems if those systems can be transparent and explain how they reach an outcome, instead of acting merely as a black box or a system for which we can only observe inputs and outputs. To that extent, transparency can enable accountability in the use of AI technology. When organizations and managers start to rely heavily on algorithms to make increasingly important decisions, they will be required to have the right explainability mechanisms in place if the results turn out to be unacceptable or difficult to understand. For instance, if IBM Watson recommends a patient treatment that seems incorrect to the physician, the physician will trust Watson’s (IBM 2018) recommendation if the system explains, in understandable terms, the factors (e.g., MRIs, scientific papers, etc.) contributing to the final outcome.

Moreover, transparency is also required to provide fair treatment to any person affected by the outcome of AI. This was the case of Sarah Wysocki (O'Neil 2016), a fifth-grade teacher at MacFarland Middle School in Washington, DC. After two years at the school, she was getting excellent reviews from the principal, as well as, the students' parents. However, at the end of the 2010–2011 academic year, Wysocki was fired because of a very bad score on her IMPACT evaluation. IMPACT is an AI tool, adopted in 2009 by the chancellor of Washington's schools with the aim of turning around the city's underperforming schools. IMPACT tries to measure the effectiveness of a given educator in teaching math and language skills. Although the initial intention of IMPACT was to minimize human bias, such as "bad teachers can seem like good ones," at the end of the day it is extremely complex to calculate the impact that one person may have on another over a year since there are many other social impacts playing in the equation. Nevertheless, when Wysocki, as well as other fired teachers, demanded details of the evaluation criteria, many school administrators were unable to provide a detailed explanation since they did not know the inner workings of IMPACT.

Transparency is challenging for organizations, as it may require revealing intellectual property by forced publishing of the AI models. Therefore, it is important to clarify situations when explanations should be given and whether they need a detailed description of AI's inner workings or rather a justification for the particular outcome (Doshi-Velez and Kortz 2017). Moreover, the new EU GDPR regulation includes the "right to explanation" by providing "meaningful information about the logic involved."

Because of the nature of current AI algorithms, especially those based on deep neural networks, it has been almost impossible to understand how AI reaches its impressive results. Nevertheless, as explainability is becoming more and more important, researchers (Binder et al. 2016) are looking at explainability mechanisms for AI algorithms, such as, layer-wise relevance propagation (LRP). These mechanisms can take an AI's outcome and work backwards through the program's neural network to reveal how a decision was made. In addition, organizations will need to hire a new type of AI professional called an AI explainer (Daugherty and Wilson 2018), whose main role will be to explain the inner workings of complex AI algorithms. In some cases, these professionals could act as algorithm forensic analysts to provide a satisfactory explanation, auditable by a competent authority, of an organization's AI system.

4.3.4. Ethics

The fourth guiding principle in the use of AI is ethics. As humans delegate certain decisions to AI systems more and more, sooner or later we will face moral dilemmas. An adapted version of the so-called trolley problem (Thomson 1976) dilemma could be found in self-driving cars. Imagine that an autonomous vehicle has a mechanical failure and is unable to stop. The AI system (acting as the driver) has two options: either the car continues and runs over and kills a family crossing the street or the car swerves, crashes into a wall and hits a bystander. Which is the moral, ethical option? Killing two people (the passenger and the bystander) or five people (the family)? In the case of a human driver, this dilemma would be solved by the judgment of the driver. However, in the case of a self-driving car using AI, ethical decisions—some of them without a right or wrong answer—should be programmed beforehand. In these situations, regulation would be needed to reflect how society (Edmond 2017) as a whole wants to deal with these ethical dilemmas.

To that extent, and inspired by Asimov's well-known Three Laws of Robotics (Asimov 1950), Oren Etzioni, professor at Washington University and chief executive of the Allen Institute for Artificial Intelligence, proposes three AI laws (Etzioni 2017) related to the ethical issues that this technology is creating. The first one is "An A.I. system must be subject to the full gamut of laws that apply to its human operator." The second is "An A.I. system must clearly disclose that it is not human." And the third is "An A.I. system cannot retain or disclose confidential information without explicit approval from the source of that information."

4.3.5. Practical Wisdom

The previous four FATE principles put the main focus on the externalities of AI on employees, customers and society at large. However, from the perspective of general management, we propose to introduce an additional principle: practical wisdom. A concept from the virtue ethics (Hursthouse and Pettigrove 2018) that is understood as the knowledge or understanding that enables its possessor to "do the right thing" in any given situation.

To that extent, ethical issues are also present when deciding spheres where organizations can apply AI technology (e.g., development of lethal weapons guided by AI). In November 2018, The World Economic Forum in their Annual Meeting of the Global Future Councils stated (Sutcliffe and Allgrove 2018): "There is a need for clearer articulation of ethical frameworks, normative standards and values-based governance models to help guide organizations in the development and use of these powerful tools in society, and to enable a human-centric approach to development that goes beyond geographic and political boundaries;" that is to say, being conscious that the focus should be the impact of AI on people, extending the concept of human rights to the digital sphere. Something that is also reflected in the 11th AI ASILOMAR principle (Future 2017): "AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity."

Therefore, organizations using AI technology in their value propositions, as well as the decision-making process, will need the new role of AI sustainers (Daugherty and Wilson 2018) to ensure that each of the AI systems satisfies its purpose of serving humans. Their overarching activities might include setting limits for AI to comply legally and ethically, managing the performance of AI and checking output quality. Future sustainer roles, such as ethics compliance manager and automation ethicist, will also include roles related to enforcing the requirement of AI algorithms to operate within human ethical and moral rules. Humans deciding which AI systems should be demoted or, on the contrary, promoted—based on their performance—will perform a role similar to HR management; however, in this case, applied to AI.

In the context of management, beyond controlling the outcomes of using AI in an organization, the use or non-use of AI must also be guided by the good judgment of the general manager, who must act on the basis of what is right for all stakeholders. This translates into managing a complex, interconnected system: the institutional configuration, the internal context, the external environment and the business, which constitutes the four genuine responsibilities of general management (Andreu and Ricart 2014). This judgment must be informed by the mission of the organization or the ultimate raison d'être of the company.

Therefore, we propose to extend the guiding principles of data usage and design algorithm with the inclusion of the practical wisdom principle related to those genuine responsibilities of the general management.

5. Conclusions

We have seen in this chapter that organizations should not consider AI as a technology in isolation. AI is becoming a reality as a consequence of living in a world with an exponential increase in digital density, where more and more data are available from the activity of companies, people and things. The increase of digital density is the underlying driving force for the process of digital transformation that business units, organizations, sectors and society as a whole are currently undergoing. This transformation touches different dimensions in the organization: the technology platform, business model and organization model. Once the organization has the required IT infrastructure in place to manage AI technology (i.e., algorithms and data), companies can leverage the benefits of AI by using it in new value propositions as part of their business models, as a combination of automation, anticipation, coordination and personalization interactions. To that extent not only is AI a powerful technology for automation purposes in many industries, but AI can also be implemented as an anticipation interaction to predict outcomes or recommend actions in a multitude of scenarios. Moreover, as value propositions are more often than not the result of the participation of many actors (companies, people, things), AI can play an important role as a coordination interaction of complex and heterogenous ecosystems. Finally, AI technology as a personalization interaction can be used to offer affordable, highly personalized products and services.

Nevertheless, in order to successfully integrate AI technology into a business model, managers should be aware of AI implications in their organization at three levels: challenges, capabilities and governance. In turn, these challenges are categorized into four categories: privacy, integration, reliability and security. Firstly, operating a business model with new value propositions based on AI requires addressing issues regarding the privacy of data derived from the human activity. Secondly, organizations need to establish the ownership of the data used to train the AI models, especially when the value proposition requires the participation and integration of many actors. Thirdly, organizations need to ensure the reliability of the outcomes of an AI model, both when it is used in new value propositions as well as in the decision-making process of the organization. Lastly, organizations must address the specific security concerns that AI poses beyond the traditional cyberattacks.

Furthermore, as data become a critical asset to create and capture value, organizations must acquire the required capabilities, both at the operational and managerial level, to successfully integrate AI in their business model. Last but not least, managers should be conscious of the potential externalities of AI on their employees, customers and society at large. To that extent, AI governance should follow the five principles of fairness, accountability, transparency, ethics and practical wisdom. Only by fully understanding the potential benefits, as well as the implications on the organization, will AI technology fulfill its promise to become a positive transformation technology of companies, sectors and society.

Table 2 summarizes the impact of AI on the business model and organizational model dimensions that general management should take into account when managing AI in a holistic way within the digital density framework.

Table 2

Impact of AI on the Business Model and Organizational Model Dimensions

Value propositions based on AI	Automation	Doing more with less resources by reducing manual steps of routine tasks.
	Anticipation	Knowing the state of the physical world in real time by describing, predicting and/or prescribing actions to be included as part of the business model.
	Coordination	Making disparate actors work together in new value propositions without the limitations of size or physical location.
	Personalization	Better meeting customer needs by using data to reflect the habits and preferences of a given customer.
AI challenges	Privacy	Establishing transparent policies about the use of personal data to preserve a relationship of trust with customers. Fulfilling compliance requirements about data privacy.
	Integration	Establishing data clauses regarding ownership and right of use of data when the value proposition derives from the participation of multiple actors.
	Reliability	Understanding the impact of the quantity and quality of the data used to train the AI algorithms, as well as the limitations and applicability of those algorithms.
	Security	Including potential risks related to adversarial attacks in the cybersecurity policies (prevention, detection and reaction) of the organization.
AI governance principles	Fairness	Identifying and eliminating beforehand noise or hidden bias in the data sources (internal and external) feeding AI algorithms.
	Accountability	Training programs for managers to assume the responsibility of the use, misuse and actions related to AI technology.
	Transparency	Excluding the use of AI systems if the results are either unacceptable or difficult to understand.
	Ethics	Ensuring that all AI systems are designed and operated to be compatible with ideas of human dignity, rights, freedoms, and cultural diversity.
	Practical wisdom	AI must be also guided by the good judgment of the general management, who must act on the basis of what is right for all stakeholders.

Source: Prepared by the authors.

In the same way that AI currently has more potential as an augmentation rather than a substitution of human activity, general managers should be aware that there is nothing artificial about AI, since machine values are derived from the human values that designed them and any bias in the input of AI systems will translate into a bias in the output. To that extent, general managers should not consider AI technology only as a tool to gain efficiency in the organization. They need to understand the consequences beyond the business model, such as the external and internal context and the institutional configuration. These implications call for general managers to develop a digital mindset to enable them to manage AI in a holistic way, preparing the IT infrastructure, and acquiring and developing the required new capabilities, creating and capturing value through new value propositions based on AI, addressing the specific new challenges that AI poses to the organization, and implementing a good AI governance based on fairness, accountability, transparency, ethics and practical wisdom principles.

References

AGRAWAL, Ajay, Gans, Joshua, and Goldfarb, Avi. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press, 2018.

ALIBABA CLOUD. "ET City Brain." Accessed December 27, 2018. https://www.alibabacloud.com/et/city.

ANDERSON, Chris. *The Long Tail: Why the Future of Business is Selling Less of More*. New York: Hyperion Books, 2006.

ANDREU, Rafael, and Ricart, Joan Enric. "The Genuine Responsibilities of the CEO." *IESE Insight* (Fourth Quarter 2014).

ANGWIN, Julia, et al. "Machine Bias." *ProPublica* (May 2016).

ASIMOV, Isaac. *I Robot*. New York: Gnome Press, 1950.

BINDER, Alexander, et al. "Layer-Wise Relevance Propagation for Deep Neural Network Architectures." *Proceeding of Information Science and Applications (ICISA)*, 2016.

BOSTROM, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

BUOLAMWINI, Joy, and Gebru, Timnit. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Research*. Conference on Fairness, Accountability, and Transparency, 2018.

CASADESUS-MASANELL, Ramon, and Ricart, Joan Enric. "How to Design a Winning Business Model." *Harvard Business Review* (January–February 2011).

DATTA, Amit, Tschantz, Michael Carl, and Datta, Anupam. "Automated Experiments on Ad Privacy Setting." *Proceedings on Privacy Enhancing Technologies*, 2015.

DAUGHERTY, Paul R., and Wilson, H. James.2018 *Human + Machine: Reimagining Work in the Age of AI*. Boston: Harvard Business Review Press, 2018.

DOMINGOS, Pedro. *The Master Algorithm*. London: Allen Lane, 2015.

DOSHI-VELEZ, Finale, and Kortz Mason. "Accountability of AI Under the Law: The Role of Explanation." Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society, working paper, 2017.

DRUCKER, Peter F. *Management: Tasks, Responsibilities, Practices*. New York: Harper & Row, 1973.

EDMOND, Awad. "Moral Machine: Perception of Moral Judgement Made by Machines," master's thesis. MIT, May 2017.

ETZIONI, Oren. "How to Regulate Artificial Intelligence." *New York Times*, September 1, 2017. Accessed December 27, 2018. https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html.

EYKHOLT, Kevin, et al. "Robust Physical-World Attacks on Deep Learning Visual Classification." *Proceeding of the Conference of Computer Vision and Pattern Recognition*, April 2018.

FAT/ML: Fairness, Accountability, and Transparency in Machine Learning. Accessed December 27, 2018. https://www.fatml.org/.

FUTURE OF LIFE INSTITUTE. "Asilomar AI Principles." 2017. Accessed December 27, 2018. https://futureoflife.org/ai-principles/.

HEALTH eHEART. Accessed December 27, 2018. https://www.health-eheartstudy.org/study.

HURSTHOUSE, Rosalind, and Pettigrove, Glen. "Virtue Ethics." *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), ed. Edward N. Zalta. Accessed January 20, 2019. https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/.

IBM WATSON. "Trust and Transparency in AI." Accessed December 27, 2018. https://www.ibm.com/watson/trust-transparency/.

INTEL. "Artificial Intelligence: The Public Opportunity." 2017. https://blogs.intel.com/policy/files/2017/10/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf.

KÁGANER, Evgeny, Zamora, Javier, and Sieber, Sandra. "The Digital Mindset: 5 Skills Every Leader Needs to Succeed in the Digital World." *IESE Insight Review*, Issue 18, (Third Quarter 2013).

KHANDANI, Amir E., Kim, Adlar J., and Lo, Andrew W. "Consumer Credit Risk Models via Machine-Learning Algorithms." *Journal of Banking & Finance*, no. 34 (2010): 2767–2787.

MALONE, Thomas. *Superminds: The Surprising Power of People and Computers Thinking Together*. New York: Little, Brown and Company, May 2018.

MICROSOFT. "FATE: Fairness, Accountability, Transparency, and Ethics in AI." Accessed December 27, 2018. https://www.microsoft.com/en-us/research/group/fate/.

O'NEIL, Cathy. *Weapons of Math Destruction*. New York: Broadway Books, 2016.

OSTERWALDER, Alexander, and Pigneur, Yves. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. New Jersey: John Wiley & Sons, 2010.

PORTER, Michael E. "How Competitive Forces Shape Strategy." *Harvard Business Review*. (March 1979).

RICART, Joan Enric. "Strategy in the 21st Century: Business Models in Action." *IESE Technical Note*, SMN-685-E, 2012.

ROSS, Casey, and Swetlitz, Ike. "IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show." *STAT*. July 25, 2018. Accessed December 27, 2018. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/.

SIEBER, Sandra, and Zamora, Javier. "The Cybersecurity Challenge in a High Digital Density World." *The European Business Review* (January–February 2019).

STEADMAN, Ian. "IBM's Watson is better at diagnosing cancer than human doctors." *Wired*. February 11, 2013. Accessed December 27, 2018. https://www.wired.co.uk/article/ibm-watson-medical-doctor.

SUTCLIFFE, Hilary, and Allgrove, Anne-Marie. "How do we build an ethical framework for the Fourth Industrial Revolution?" *World Economic Forum*. November 7, 2018. Accessed December 1, 2018. https://www.weforum.org/agenda/2018/11/ethical-framework-fourth-industrial-revolution/.

THOMSON, Judith, J. "Killing, Letting Die, and the Trolley Problem." *Monist: An International Quarterly Journal of General Philosophical Inquiry*, vol. 59, 1976.

WING, Jeannette, M. "Data for Good: FATES, Elaborated." Data Science Institute, Columbia University, January 23, 2018. Accessed December 27, 2018. https://datascience.columbia.edu/FATES-Elaborated.

WRZESZCZYNSKI, Kazimierz, O., et al. "Comparing sequencing assays and human-machine analyses in actionable genomics for glioblastoma." *Neurology Genetics* (August 2017).

ZAMORA, Javier. "Making Better Decisions Using Big Data." *Harvard Deusto Business Review* (May 2016), ref. 016197-E.

ZAMORA, Javier. "Programming Business Models Through Digital Density," *IESE Insight* (Second Quarter 2017).

ZAMORA, Javier, and Herrera P. "How prepared is your business to make the most of AI?" *IESE Business School Insight*, no. 151, (Winter 2018).

ZAMORA, Javier, Tatarinov, Katherine, and Sieber, Sandra. "The Centrifugal and Centripetal Forces Affecting the Digital Transformation of Industries." *Harvard Deusto Business Review*, no. 279 (June 2018), ref. 018189.