Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance

Michael Hilb¹

Published online: 21 July 2020
© The Author(s) 2020

Abstract

The article explores the impact of the ongoing progress and adaptation of artificial intelligence on the practice of the corporate governance. It applies three lenses to artificial governance—the business, technology and society lenses—to assess the desirability, feasibility and responsibility of automating board-level decision-making to ensure effective corporate governance. Based on an assessment of the potential and limitations of human and machine learning for effective board-level decision-making, the article proposes five scenarios of artificial governance, i.e. assisted, augmented, amplified, autonomous and autopoietic intelligence, that are likely to shape the governance of organizations today, tomorrow and beyond. It discusses the implications of both the governance *of* and the governance *with* artificial intelligence in the three horizons and concludes with an appeal to board members to take an active role in understanding, imagining and shaping the future of artificial governance.

Keywords Artificial governance · Corporate governance · Artificial intelligence · Machine learning · Decision making

1 Introduction

Although artificial intelligence (AI) is now at the top of the agenda for many business leaders (Davenport and Ronanki 2018), it is not a new term—it was originally coined in the 19050s (Russell and Norvig 2016). Its importance for corporate management and governance, however, has long been ignored, as Peter Drucker’s article “The Manager and the Moron” stipulates (Drucker 1967, p. 49): “The computer does not make decisions, it only executes commands. It’s a total moron.”

How times have changed. AI is now widely considered a “general purpose technology” (Mantas 2019, p. 42), by many even seen as a “general solution technology,” i.e. the solution to any managerial, commercial, or even societal problem.

✉ Michael Hilb
michael.hilb@unifr.ch

¹ University of Fribourg, Fribourg, Switzerland

What is still receiving less attention in the current state of euphoria is the impact of AI on the concept of the corporation and its governance itself (Libert et al. 2017). This article aims to close this gap by proposing a scenario framework to assess the impact of AI on corporate governance practice.

For the purpose of this article, we define corporate governance as “the system by which companies are directed and controlled” (Cadbury 1992, p. 15). The system can be defined as the composite of “ownership, boards, incentives, company law, and other mechanisms” (Thomsen 2008, p. 15). We define artificial intelligence (AI) as the “activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment” (Nilsson 2010, p. 13). Hence, we shall focus in this article on the impact of *intelligent machines* on the activity of *decision-making* by the entity of the *board of directors* (BoD) related to the *control and direction* of the corporation.

Given the multifaceted nature of corporate governance, we propose an integrated perspective combining the business, i.e. the definition of the realm of *desirability* and the technology, i.e. the definition of the realm of *feasibility*, with the legal and ethical perspectives. In line with the neo-institutional view that ethical considerations eventually evolve into legally binding rules (Scott 1995), we merge the latter two into one perspective, which defines the realm of *responsibility*.

2 The realm of desirability: the business perspective

To better understand the potential contribution of AI to the decision making of the BoD, we need to better understand the anatomy of this process. To this end, we shall first define the key functions and archetypical decisions of corporate boards. In a second step, we will characterize the level of predictability of those decisions given the pivotal role of predictions in AI (Agrawal et al. 2018).

2.1 Defining a taxonomy of board decisions

Although national laws differ in terms of the roles and responsibilities assigned to directors, Cossin and Metayer (2014) identify three generic key roles of BoDs across all jurisdictions: supervisor, co-creator and supporter, hereby extending the traditional dualistic perspective of direction and control. We shall base our role definition on Cossin and Metayer (2014) and label them as follows:

- *Co-direction* The BoD is responsible for strategic leadership, for developing the corporate strategy together with the top management team (TMT) and for ensuring proper strategy implementation by setting objectives.
- *Control* Another key responsibility of the BoD is to control the TMT and to ensure full compliance with the law, accounting codes, and the company’s statutory rules, particularly with regard to the company’s finances and risk management.

- *Coaching* The BoD is also responsible for appointing and coaching the TMT to ensure effective leadership.

As we focus on the impact of AI on the BoD's decision making, we need to identify the key decision types a BoD usually deals with (there are other non-routine tasks such as crisis management or communication which we ignore for the purpose of this article):

- *Co-direction* (a) decision on innovation; (b) decision on collaboration; (c) decision on optimization; (d) decision on transformation; (e) decision on diversification/concentration; (f) decision on internationalization.
- *Control* (a) decision on target achievements; (b) decision on meeting accounting standards; (c) decision on legal compliance; (d) decision on ethical compliance.
- *Coaching* (a) decision on executive appointments; (b) decision on executive development; (c) decision on executive compensation; (d) decision on board composition.

To determine which types of decisions benefit most from being driven by AI, it is important to recognize an inherent feature of decision making. Decision making is always about consciously choosing between two or more options. The options can be either binary, e.g. yes or no, or multifaceted, e.g. options 1, 2 and 3. The choice always depends on the criteria chosen. An informed decision usually follows a similar pattern as outlined by Still et al. (1958). They distinguish between three phases, i.e. conceptualization, information, and prediction (Fig. 1). We apply this taxonomy and add a second level to distinguish the sub-processes even further (For simplification, we do not consider the decision feedback loop usually associated with decisions at this point).

- *Decision sensing* At the beginning of every decision is the need to change or validate the course of action and to make a decision. This requires a certain ability to comprehend the context.
- *Decision framing* The conceptualization of a decision is the key to ensuring that all parties involved agree and have a clear understanding of the results.
- *Information collection* Information is a key component of any decision. Gathering relevant information is of central importance and requires experience.
- *Information selection* Just as important as the gathering of information is the selection of the relevant information needed to reach a conclusion.
- *Option identification* As a decision is about options, the possible outcomes need to be predicted first.

Fig. 1 The anatomy of a decision. (adapted from Still et al. 1958)

Fig. 2 The four decision types (adapted from Stacey 1992)

- *Option assessment* The final decision depends on the valuation of the option as compared to the valuation of the alternative options.

2.2 Proposing predictability levels of board decisions

In order to assess whether it is desirable to use AI in business decisions, we propose to apply the taxonomy of the different decision types according to Stacey (1992) which distinguishes between four types of decisions depending on the degree of certainty and agreement (Fig. 2):

- *Common decisions*¹ Certain decisions are considered to be fairly straightforward as the outcome is certain, and all decision-makers are in full agreement.
- *Complicated decisions* The second type of decision is placed in a multi-optional context, which usually requires different points of view.
- *Complex decisions* These types of decisions are made in a context that is either totally uncertain or leads to significant disagreement.
- *Chaotic decisions* After all, there are decisions that have to be made in a completely fluid environment, which, by nature, leads to different points of view.

How can this logic be applied to board-level decision making? The table below summarizes which decisions are common, complicated, complex, and chaotic from a governance perspective (Table 1).

We have seen that board decisions are of varied natures when it comes to certainty and agreement. How can AI be utilized to facilitate or even drive board decisions?

¹ Stacey (1992) calls this decision "simple". We will instead call it "common" to better characterize its nature as routine activities.

Table 1 Corporate governance decisions by decision type

	Conceptualisation		Information		Prediction		Option assessment
	Decision sensing	Decision framing	Information collection	Information selection	Option identification		
Co-direction							
Innovation	Chaotic	Complex	Complex	Complicated	Complicated	Complicated	Complicated
Collaboration	Complex	Complex	Complex	Complex	Complicated	Complicated	Complex
Optimization	Complicated	Complicated	Complicated	Common	Complicated	Common	Common
Transformation	Complex	Complex	Complicated	Complex	Complex	Complex	Complicated
Diversification	Complex	Complex	Complicated	Complex	Complex	Complex	Complicated
Internationalization	Complex	Complicated	Complicated	Complex	Complex	Complex	Complicated
Control							
Target achievement	Common	Common	Complicated	Common	Common	Common	Common
Accounting standards	Common	Common	Complicated	Common	Common	Common	Common
Legal compliance	Complicated	Complicated	Complicated	Common	Complicated	Common	Common
Ethical compliance	Complex	Complicated	Complex	Complex	Complex	Complex	Complicated
Coaching							
Executive appointments	Complex	Complicated	Complex	Complex	Complex	Complex	Complicated
Executive development	Complex	Complex	Complex	Complex	Complex	Complicated	Complicated
Executive compensation	Complicated	Complicated	Complicated	Common	Complicated	Common	Common
Board compensation	Complicated	Complicated	Complicated	Common	Complicated	Complicated	Common

3 The realm of feasibility: the technology perspective

In order to better understand the potential contribution of AI on decision making at the board level, it is indispensable to decipher the black box of AI and discuss its different types and evolution (Nilsson 2010) as well as its advantages and shortcomings as compared to human intelligence.

3.1 Understanding different approaches to artificial intelligence

AI is a technology in the making, as explained above. Although we are currently seeing important advances in AI, this is not the first and will not be the last wave of development. It is therefore important to look at AI from a dynamic perspective, i.e. consider the “trajectory of AI” (Armour and Eidenmueller 2019). As Tegmark (2017) notes, we are currently in the early phase of AI development.

At the highest level, we can distinguish between rule-based systems and machine learning (ML). The former requires that humans fully understand a given context and define the rules that the machine should execute, while the latter enables the machine to learn and derive conclusions based on a set of data and learning algorithms without requiring context understanding (in the end, ML could also be considered rule-based, if only for the rules underlying the learning algorithms).

Within ML, deep learning (DL) is currently the most popular approach. It owes its name to its architecture of “deep (or multi-layered) artificial neural networks—software that roughly emulates the way neurons operate in the brain” (Ford 2018, p. 10). Earlier, more traditional approaches to ML still require the extraction of features by humans. Within DL, three approaches are usually distinguished, i.e. supervised (SL), reinforcement (RL), and unsupervised learning (UL). Supervised learning, currently the most commonly used approach, requires well-structured and labeled training data to train the algorithms to improve AI-driven applications such as image recognition or translation. In contrast, the philosophy behind reinforcement learning is best described by trial and error, which is often used in board game simulations or. The challenge of RL lies in the large number of trial rounds required to achieve good results. Currently, the most promising, but also most challenging approach is unsupervised learning, where the algorithms are designed to “learn directly from unstructured data coming from their environments” (Ford 2018, p. 12).

All these approaches assume a separation of the machine from the mind. The next wave of AI could overcome this separation and be based on the assumption that machines and minds can be connected. Whether we will see the emergence of “neuromorphic chips” (Marsh 2019) or other ways to connect the mind to the machine, this type of AI is likely to unleash new potentials. We will call this futuristic category Mind Machine Learning (MML) (Fig. 3).

In this article, we focus on the impact of the technologies currently used and developed, i.e. SL, RL and UL, and the technologies currently being researched, i.e. MML. For simplicity, we will continue to use the term AI to refer to implications common to all the approaches discussed.

Fig. 3 The taxonomy of artificial intelligence and analytics. (Partly adapted from FSB 2017)

3.2 Understanding the power and limitations of different approaches to artificial intelligence

To determine the strengths of the different approaches to artificial intelligence, it is important to be clear about the scale, i.e. human intelligence, and how the process of intelligence development, i.e. learning, is compared between the mind and the machine.

To address the latter question, we will briefly compare human and machine learning cycles based on the decision model presented above. Both approaches assume that decisions are based on predictions of possible outcomes, whereas prediction is defined as “the process of filling in missing information. Prediction takes information you have, often called “data,” and uses it to generate information you don’t have” (Agrawal et al. 2018, p. 24). In both approaches, predictions are based on input data.

However, in contrast to human learning, ML assumes three types of data, which play a different role in SL, RL and UL: Input data (direct input to the algorithm leading to predictions), training data (used primarily to generate the algorithm), and feedback data (used to improve the algorithm over time) (Agrawal et al. 2018). While all three approaches to ML rely on input data, training data is crucial for SL, while feedback data is crucial for RL and UL (Fig. 4).

This brings us back to the first question: the power of AI. Here, it is important to discuss which learning approach is best suited for which decision type, as discussed earlier. While SL can be used for common decisions, it becomes much less effective for other decision types, i.e. complicated, complex or chaotic, given the need for relevant training data. RL, on the other hand, can be effective in automating complicated decisions based on past routines. However, because it relies heavily on trial and error and thus on feedback data, it is not very effective in handling complex or even chaotic decisions. For complex decisions, UL could provide clues, while for chaotic decisions, it will be even more difficult to rely on any known ML approach. Any machine that would succeed in mastering chaotic decisions would pass the final Turing test, i.e. possess artificial general intelligence as proposed by Alan Turing in his 1950 paper “Computing Machinery and Intelligence” (Levesque 2017). Or to put it differently, the current stage of AI faces a number of serious obstacles as

Fig. 4 Human and machine learning cycles (partly adapted from Still et al. 1958; Agrawal et al. 2018)

summarized by Marcus (2018): There are a number of problems arising from the lack of transparency or transferability, the inherent inability to distinguish between causality and correlation, and inefficiency in terms of the amount of data required for valid predictions.

Here comes the hope for alternative approaches to ML, such as MML, as indicated above. However, the true potential and limitations of this type of approach have yet to be seen. The question of whether and when artificial general intelligence will become a reality is still up for debate.

4 The realm of responsibility: the society perspective

“The saddest aspect of life right now is that science gathers knowledge faster than society gathers wisdom.”

Isaac Asimov as cited by Tegmark (2017, p. 316).

As business, society and politics are in the middle of the process of adopting AI, the public discourse on the proper regulatory frameworks is on-going. As is often the case with new technologies, technological development and business applications run faster than regulation. In this sense, today’s legal and ethical issues are very similar to those that were raised in the medical field in the 1700s (Gasser and Schmitt 2019).

Some of the most important legal and ethical issues relevant to corporate governance are now emerging. We will focus on five of each in the following sections.

4.1 Legal considerations: complying with AI regulations

The BoD will be guided by five key legal considerations in its decision to incorporate and integrate AI into its processes and procedures.

- **Accountability** Since accountability is at the core of corporate governance, the impact of AI on this principle is central. As outlined by Wooldridge and Micklethwait (2003), the concept of the company was a revolutionary idea at the time of its

creation, as the creation of a legal entity with limited liability allowed companies to take more risks than would be possible for a single person. In order to avoid abuse of the limited liability construct, the management and supervision of companies were entrusted to natural persons who were responsible for the performance of their tasks according to clearly defined criteria. The concept of delegation is central. While delegation generally applies between natural persons, e.g. the BoD can delegate tasks to the TMT or another legal entity, e.g. an audit firm, the core tasks of a BoD member, i.e. the direction and control of a company, cannot be delegated. For the time being, this also applies to the delegation to machines. Thus, even if a BoD were to automate the entire decision-making process using AI, BoD members would still remain accountable under the current regime.

- *Liability* The legal concept of liability is linked to the first point. Any natural person who exercises a fiduciary duty is liable for any failure to act effectively in his or her role as director of a company. Who would be liable if an error in an algorithm led to wrong decisions? The user of the algorithm, its developer or its vendor? What happens if a company develops the algorithm internally? Will there be insurance to cover the risks? (Armour and Eidenmueller 2019). These are critical legal issues that have not yet been fully tested in court.
- *Business judgement* The business judgment rule states that any key decision taken at the board level must be based on the best available information and that the BoD must document the decision accordingly. This has two implications for AI and corporate governance: On the one hand, companies could be forced to resort to AI if it promises better results than those from people with limited rationality. On the other hand, the black boxes behind many AI applications would need to be decrypted in the event of legal disputes.
- *Data protection* Since the effectiveness of AI depends very much on the availability of data as outlined above, the regulation of data protection and data access is crucial to ensure that companies can protect sensitive data needed for strategic decisions. At the same time, companies often need to have access to publicly available data to ensure that all relevant data points can be considered to optimize AI. This dual challenge poses major risks for companies to fully-automate decision making at the board level.
- *Regime heterogeneity* Since the business rules for different markets are still tied to nation-states, companies engaged in international business are exposed to a number of regulatory regimes that take very different positions on AI-related issues (Hilb 2017). Boards must understand the opportunities and risks of such an engagement to ensure that international data flows can be maintained to allow for the best possible decisions to be made.

4.2 Ethical considerations: anticipating societal AI expectations

In addition to a number of legal questions, there are even more serious ethical considerations to take into account (Simonite 2018). The five key ethical questions can be summarized as follows:

- *Bias in and by AI* Bias plays an important role in the human decision-making process. This also applies to ML, where machines imitate human solution finding by relying on input data that may be biased (Rosso 2018). It is crucial to recognize these prejudices and to have the courage to correct them. At the same time, the results of AI will influence human behavior and may cause other distortions.
- *Distribution of wealth* A second ethical issue that BoDs need to address is data ownership since data is the most important asset class in the “intangible economy” (Haskel and Westlake 2018). The ethical question of who is to benefit from the economic value from data will be one of the most important debates in the future, which boards cannot ignore. They must be prepared to adapt their approach to AI as society will ask for a fair distribution of the benefits of AI (Altman 2015).
- *Monopolization of intelligence* In the context of the former consideration, the monopolization of intelligence capabilities, the key component of effective AI, is a key factor in the development of the AI system. Since knowledge means power, the ability to create intelligence will become the new gold mill. With power comes responsibility. This applies to the BoD as the higher-level management body responsible for ensuring corporate responsibility. In particular, the BoD is responsible for ensuring that none of the AI technologies are misused.
- *Values* Another central concern in the use of AI in corporate governance is the question of what the moral basis of any decision should be. What should the underlying principles be, and how can such principles be integrated into AI? These simple questions are linked to a deeper philosophical debate about whether there is a single truth or whether the underlying assumptions need to be made explicit. This is a central question, especially for DL, since any learning can only take into account the morality applied in the past. As a consequence, AI assumes an instrumental rather than a value-based view of morality. Since values are central to any successful companies, such an instrumental view of morality can lead to unintended consequences and conflicts that should be avoided at all costs.
- *Free will* Finally, there are legitimate concerns that autonomous decision-making systems may limit the autonomy and free will of individuals or companies, one of the key concepts of civilization. Especially as an independent board member, free will is a key characteristic that must be preserved and cultivated.

5 The realm of sustainability: the integrated perspective

When we combine the areas of what is desirable, feasible and responsible, we enter the realm of what is sustainable for AI in corporate governance. The above discussion has shown that the application of AI in corporate governance is not only complex but also dynamic. Therefore, the discussion must be about possible scenarios and not about a single point of view. Furthermore, the above discussion has brought to the surface the prevailing duality in all dimensions of artificial intelligence: mind and machine. There are always two ways to look at dualities: You can consider them as competing or complementary. While there is a natural

competition between the human mind and the machine for efficiency and effectiveness of intelligence, the two approaches to intelligence are also highly complementary. Therefore, the highest level of intelligence must be synergic, i.e. a state in which human and machine intelligence function and are superior in combination (Hilb 2019). What does this synergic intelligence look like?

We will distinguish between five scenarios of synergic intelligence [partly adapted from Nalder (2017) and Armour and Eidenmueller (2019)]:

- *Assisted intelligence* In the case of assisted intelligence, humans are still the decision-makers who rely on selective decision support from AI-driven applications such as translation or speech recognition. These approaches are generally accepted and even appreciated by society and are usually well regulated.
- *Augmented intelligence* While humans remain the clear decision-maker when it comes to applying augmented intelligence, the AI-based solutions used are more sophisticated and allow the decision-maker to use the technology in a way that surpasses human intelligence, e.g. by identifying outliers in large amounts of data or automated reporting. The regulation of augmented intelligence is at the top of the agenda today as more and more implications become visible and call for regulation.
- *Amplified intelligence* The use of amplified intelligence requires joint decision-making by man and machine, i.e. the machine can make a recommendation that must be approved by man, who is able to provide additional inputs, e.g. in the case of complex expert recommendations. A coexistence between mind and machine is nowadays neither socially acceptable nor provided for in any legal regime. Therefore, the social debate still needs to take full shape.
- *Autonomous intelligence* With autonomous intelligence, the machine can make decisions independently and operate within a predefined range without constant decision inputs. Examples of this are self-regulating control mechanisms or highly developed robots. Social and regulatory debates on how to deal with autonomous intelligence have begun, but have not yet led to a generally accepted consensus, as some of the debates, e.g. on accountability and liability, challenge the basic assumptions of today's regulatory framework.
- *Autopoietic intelligence* The application of autopoietic intelligence is based on an artificial entity that is not only capable of making independent decisions within a certain area, but is also able to develop and expand this area over time. It marginalizes the necessity and influence of human decision making. Examples of the application of this type of intelligence can be found in science fiction literature up to this point. As a result, a substantive societal debate on how to apply autopoietic intelligence has yet to start.

How do human and machine intelligence interact? At the heart of the concept of synergic intelligence is the conviction that the result of such combinations yields higher results than just their addition. This is illustrated by the graph below, which shows the intelligence efficiency and intelligence effectiveness

Fig. 5 The five scenarios of synergic intelligence by intelligence efficiency and effectiveness (the five scenarios are partly adapted from Nalder 2017; Armour and Eidenmueller 2019)

functions. While the former assumes that the machine replaces humans only to achieve efficiency with the same result, the intelligence effectiveness function shows how better decisions can be made by combining both approaches under the assumption of limited human intelligence. True synergic intelligence is, therefore the result of the intelligence effectiveness function (Fig. 5).

How can these findings be applied to AI in corporate governance? Let us analyze their impact on the three roles of a board of directors as defined earlier for each of the five scenarios:

5.1 Assisted intelligence: making the board more efficient in governance

Applied to the context of corporate governance, the application of assisted intelligence does not influence the basic principles of board practice but rather provides board members with additional tools to support their decision making. The human decision-maker, i.e. the board member, remains at the center of power. The implications along the three key functions of the board of directors are as follows:

- *Co-direction* The use of AI-driven tools to better deal with market and operational data provides valuable inputs for board members involved in strategic decision making.
- *Control* At the same time, the use of AI allows to further automate the corporate consolidation and reporting processes providing the board with real-time data increasing the overall transparency of corporate affairs.

- *Coaching* The coaching-related impact of AI technology is marginal, as the use of large amounts of data to foster interpersonal relationships between members of the BoD and TMT is marginal, given the people-centric nature of this partnership.

5.2 Augmented intelligence: making the board more effective in governance

The use of augmented intelligence neither calls into question the fundamental nature of corporate governance, nor does it deprive board members of power. Rather, it promises to make decision-making not only more efficient but also more effective, as intelligent technologies are used to deliver better results. Along with the three key functions of the board of directors, the following implications arise:

- *Co-direction* The key contribution of AI lies in the use of predictive models, which help to develop more valid scenarios and superior simulations that improve decision making in strategic board decisions, both in terms of investment and people.
- *Control* The use of improved forecasting capabilities in the supervisory boards will help to change the nature of controls from predominantly past-oriented to future-oriented ones. This change will lead to a fundamental shift in the role and influence of the board of directors in ensuring control of the company.
- *Coaching* Predictive insights also help to develop key people and make compensation decisions more data-driven.

5.3 Amplified intelligence: making the board and machine co-govern

The trend towards amplified intelligence is beginning to challenge some of the basic business and legal assumptions of corporate governance. As amplified intelligence assumes an evenly distributed contribution and accountability of man and machine, the human-centered approach to governance is being challenged. There is a need to discuss how to deal with the basic assumption of corporate law, i.e. personal accountability for corporate liabilities, when effective decision making is shared. Who should be held accountable for what? In this scenario, the implications along the three dimensions are as follows:

- *Co-direction* Some goal-setting decisions are fully automated, while others remain in the hands of people. The distinction between machine-driven and human-driven decisions is likely to be based on the respective ability, but also on the assumed effect.
- *Control* The same logic applies to the control function of the board. Many compliance tasks are fully automated, making them more reliable and secure against misuse and mishandling. However, some of the human-centric compliance issues still need to be managed by humans. Similar segregation of duties is assumed

when it comes to dealing with risk and uncertainty issues. While most of the former decisions are made by machines, human involvement increases the validity of decisions related to uncertainty.

- *Coaching* In this scenario, coaching becomes an extended meaning, since coaching is not limited to dealing with people, i.e. other board members or managers, but also with machines. They must also be trained and maintained. Both are board duties.

5.4 Autonomous intelligence: making the corporation self-govern

In the autonomous intelligence scenario, all or some board members are replaced by machines, either by a governance robot (a robot representing the board of directors) or by robo-directors (various robots that simulate contributions from individual directors), depending on how much the legal framework changes. At the same time, it is still assumed that people will retain responsibility for determining the scope of corporate governance. There is already a real case study in Hong Kong where Vital, an AI system, is one of many board members in an investment company (Burrige 2017). As highlighted by Armour and Eidenmueller (2019), this scenario is most likely realized in subsidiaries first. The effects on the three key roles in this scenario are as follows:

- *Co-direction* The steering robot or robo-directors make strategic decisions within a predefined area independent of case-related instructions. However, their algorithms must be certified for the tasks assigned to fulfil the legal obligations foreseen for the governance of the institutions.
- *Control* The same applies to their role in ensuring adequate control. Since rules, regulations and standards are central to this area, the certification of governance robots or robot directors can even be regulated. This is important when relying on such mechanisms for cross-company cooperation, for instance.
- *Coaching* In this scenario, coaching becomes synonymous with machine development and maintenance. People are likely to play a role in this as machines that coach other machines.

5.5 Autopoietic intelligence: making corporate governance self-evolve

In a scenario where autopoietic intelligence dominates, the governance of a business entity is not only fully automated to ensure proper governance within a certain framework, e.g. the representation of clearly defined shareholder interests, but also capable of driving its future development and interacting proactively with other business entities. In this scenario, the impact on the three key roles is as follows:

- *Co-direction* Both the setting of the agenda and the strategic decisions themselves are fully automated and fully comply with the requirements for good business judgment. Human intervention is no longer necessary.

- *Control* A fully automated feedback system ensures that the goals set are constantly monitored, measured but also challenged. The co-direction and control functions are therefore closely linked.
- *Coaching* The self-development of the system also ensures continuous improvement through feedback within and between the units involved. An effective coaching function based on this understanding will be the main difference between better and worse functioning governance systems.

6 Implications

What do these scenarios mean for the board members? How should they begin to think about the opportunities, but also the risks that arise from introducing AI in their boardrooms? What are the key questions to ask?

In line with the dynamic perspective on AI as introduced previously, we will apply the innovation horizon model proposed by Baghai et al. (2000) and adapt it to AI-driven governance. They recommend that companies simultaneously focus on continuous innovation to improve a company's existing business model (Horizon 1), extend a company's existing business model to new customers, segments or markets (Horizon 2) and create new businesses to take advantage of future disruptive opportunities (Horizon 3). Furthermore, we will highlight the implications for both governance *of* artificial intelligence and governance *with* artificial intelligence.

6.1 Horizon 1: exploiting the opportunities offered by current AI to improve corporate governance today

Artificial governance in Horizon 1 is very much driven by advances in supervised and, to a lesser extent, reinforcement learning, which mainly leads to assisted and augmented intelligence. In both cases, correct and sufficient data is key to performance, either as training or feedback data.

The governance of data is the key focus when it comes to governance *of* AI in Horizon 1. Companies increasingly understand the strategic value of data as training, but even more importantly, as feedback data (Iansiti and Lakhani 2020). For many companies, data is becoming the most valuable asset (Mayer-Schönberger and Ramge 2018) in line with “the rise of the intangible economy” (Haskel and Westlake 2018). It is essential to understand the economic characteristics of data, in particular, that the data is not competitive, has a high option value, often has high up-front and low marginal costs, and requires complementary investments such as technology and qualified staff (Coyle, Diepeveen, Wdowin, Kay, and Tennison 2020). Therefore, the governance of these key assets is of strategic importance and needs to be addressed at the board level (Hilb 2019). At the same time, data-related issues pose major risks and increasingly determine corporate culture and the relationship between companies and society (Mantas 2019).

At the same time, it is important that the board recognizes that AI does not only affect the business but also the board itself, i.e. the governance *with* AI. This

awareness is necessary to make the BoD more data-driven (Libert et al. 2017) to take full advantage of the benefits of what Davenport and Ronanki (2018) call the “future cognitive company”. A key focus in Horizon 1 is on reporting and controlling, as AI enables sample-based audits, now data and stronger predictive power of data (Mantas 2019). This will eventually lead to a reduction in the cost of internal control, but also to increased supervision and liability risks beyond the company (Armour and Eidenmueller 2019).

6.2 Horizon 2: exploring the opportunities of future AI to improve corporate governance tomorrow

The impact of artificial intelligence on corporate governance will reach a new level in Horizon 2, mainly through the development of reinforcement and unsupervised learning, which will extend the impact of artificial intelligence lending support to augmented, amplified and autonomous intelligence. For the first time, the vision of a self-propelled organization is more than just science fiction. Systems will be able to exceed human learning capabilities not only in narrow areas but also in broader, interrelated decisions, as described above.

In such a context, the algorithm becomes the focus of attention of the governance of AI. Those organizations that are able to develop and apply an algorithm that enables reinforcement and, more importantly, unsupervised learning, will be best positioned to benefit from it. Since no single organization will be able to develop or use such algorithms in isolation to achieve their full potential, collaboration between organizations in ecosystems will be critical (Hilb 2017). This results in some legal issues, such as liability for the algorithm, i.e. should those who use or build it be held liable for its result, antitrust law, i.e. can a superior algorithm lead to anti-competitive behavior, especially when companies join forces to develop and use it, or related regulatory issues such as tax law, i.e. how should the value-added of algorithms be taxed? At the same time, in-house procedures and processes must be put in place to ensure that all these questions are addressed properly, or as Armour and Eidenmueller (2019, p. 36) put it: “(T)raditional agency problems in corporations will increasingly lose importance as an analytical issue: algorithms optimize a given goal function. At the same time, new oversight challenges and, as a consequence, ‘strategic’ agency costs are created at the top of a firm’s hierarchy.”

When it comes to the governance *with* AI, the focus will shift from just testing to applying artificial intelligence as a legally binding requirement. As artificial intelligence will be more reliable and also applicable for setting objectives, AI-led decision making will in many cases be the most reliable way to arrive at a decision that meets the criteria of “good business judgment,” as provided for as a key principle in most company law regulations. Therefore, the direction of an enterprise is increasingly determined by the AI setting the target itself, the control of which is also determined by the AI as described in Horizon 1, thus closing the loop.

6.3 Horizon 3: shaping a new corporate governance model by taking advantage of the disruptive power of futuristic AI

Horizon 3 of artificial governance still holds many unknowns, since we do not know how AI technology will evolve, nor what rules society will impose on those who develop and use it. What we do know, however, is the direction in which we are heading. There will be approaches to artificial intelligence that go beyond neural networks and include direct connections between the mind and the machine (Marsh 2019). This approach is likely to be disruptive in nature as defined by Christensen (1997), i.e. to achieve a new level of efficiency and effectiveness, whereby old approaches eventually become obsolete. These approaches are likely to bring us closer to what is known as general intelligence (Marsh 2019) that is, what we call autonomous and autopoietic intelligence.

For the governance *of* AI, this means that the focus shifts to the interface between mind and machine and how to draw from these links. It goes without saying that this level of linkage will raise a whole new set of legal and ethical questions that touch the very substance of human will and self-determination. Since the technological possibilities of organizations to influence human thought will be greater and more difficult to control, their regulation will be highly controversial but also influential. At the same time, organizations will probably become self-organizing systems in the truest sense of the word, i.e. they will act without any human intervention.

This will also shift the governance *with* AI to new levels. AI will enable the organization to control and direct itself more efficiently and effectively than any human being could. There will, therefore, be economic pressure from the capital market to adapt to AI-driven forms of governance. This model is likely to challenge the basic tenant of today's corporate law: The personal liability of the directors of limited liability companies (Moeslein 2018) (Table 2).

7 Conclusions

“Success in creating effective AI, could be the biggest event in the history of our civilization. Or the worst. We just don’t know.”

Stephen Hawking as cited by Girasa (2020, p. 64).

While the business appetite for AI is clear and the advances in technology are certain, it will ultimately be the social dialogue that will be critical. In this respect, companies will have to prove that they are aware of their responsibility in dealing with AI in order to win the trust of society. At the same time, AI will have a profound impact on corporate governance, as it will enable the introduction of a whole range of new governance mechanisms and systems. The result could be a refined concept of the limited-liability company, the capital market and, consequently, capitalism. Today's boards of directors can play a central role in this process if they are willing and able to take the driver's seat.

Table 2 The three horizons of artificial governance

	Horizon 1 of artificial governance	Horizon 2 of artificial governance	Horizon 3 of artificial governance
Characteristics			
Learning focus	Supervised and reinforcement machine learning	Reinforcement and unsupervised machine learning	Mind machine learning
Intelligence focus	Assisted and augmented intelligence	Augmented, amplified and autonomous intelligence	Autonomous and autopoietic intelligence
Implications on the governance of AI			
Asset focus	Data	Algorithm	Mind machine interface
Unit focus	Corporation	Ecosystem	Self organization system
Implications on the governance with AI			
Mechanism focus	Control	Direction	Self-control and self-direction
Attention focus	Awareness for artificial governance	Application of artificial governance	Adaptation to artificial governance

Acknowledgements Open access funding provided by University of Fribourg.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agrawal, A., Gans, J. S., & Goldfarb, A. (2018). *Prediction machines: the simple economics of artificial intelligence* (Vol. Spring). Boston: Harvard Business Press.
- Altman, R. (2015). Distribute AI benefits fairly. *Nature*, 521, 417–418.
- Armour, J., & Eidenmueller, H. (2019). *Self-driving corporations?* ECGI Working Paper Series in Law.
- Baghai, M., Coley, S., White, D., & Coley, S. (2000). *The alchemy of growth: practical insights for building the enduring enterprise*. New York: Basic Books.
- Burridge, N. (2017). *AI takes its place in the boardroom*. Tokyo: Nikkai Asian Review.
- Cadbury, A. (1992). *Code of best practice: report of the committee on the financial aspects of corporate governance*. London: Gee and Co.
- Christensen, C. M. (1997). *The innovator's dilemma: when new technologies cause great firms to fail*. Boston: Harvard Business Review Press.
- Cossin, D., & Metayer, E. (2014). *How strategic is your board?*. Cambridge: MIT Sloan Business Review.
- Coyle, D., Diepeveen, S., Wdowin, J., Kay, L., & Tennison, J. (2020). *The value of data*. Cambridge: Bannett Institute of Public Policy, Cambridge University.
- Davenport, T. H., & Ronanki, R. (2018). *Artificial intelligence for the real world*. Massachusetts: Harvard Business Review.
- Drucker, P. F. (1967). The manager and the moron. *McKinsey Quarterly*, 3(4), 42–52.
- Ford, M. (2018). *Architects of intelligence: the truth about AI from the people building it*. Birmingham: Packt Publishing.
- FSB. (2017). *Artificial intelligence and machine learning in financial services—market developments and financial stability implications*. Basel: Financial Stability Board.
- Gasser, U., & Schmitt, C. (2019). The role of professional norms in the governance of artificial intelligence. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics and AI (forthcoming)*. Oxford: Oxford University Press.
- Girasa, R. (2020). *Artificial intelligence as a disruptive technology*. Cham: Palgrave Macmillan.
- Haskel, J., & Westlake, S. (2018). *Capitalism without capital*. Princeton: Princeton University Press.
- Hilb, M. (2017). The co-evolution of digital ecosystems. In M. Hilb (Ed.), *Governance of digitalization* (pp. 108–114). Bern: Haupt.
- Hilb, M. (2019). *Unlocking the board's data value challenge*. Directorship (pp. 60–61).
- Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI*. Boston: Harvard Business Review Press.
- Levesque, H. (2017). *Common sense, the turing test, and the quest for real AI*. Cambridge MA: MIT Press.
- Libert, B., Beck, M., & Bonchek, M. (2017). AI in the boardroom: the next realm of corporate governance. MIT Sloan Management Review Blog.
- Mantas, J. B. (2019). *Intelligent approaches to AI*. Directorship (pp. 40–45).
- Marcus, G. (2018). *Deep learning: a critical appraisal*. New York: New York University.
- Marsh, H. (2019). *Can man ever build a mind?*. London: Financial Times.
- Mayer-Schönberger, V., & Ramge, T. (2018). *Reinventing capitalism in the age of Big Data*. London: John Murray.

- Moeslein, F. (2018). Robots in the boardroom: artificial intelligence and corporate law. In W. Barfield & U. Pagallo (Eds.), *Research handbook on the law of artificial intelligence* (pp. 649–670). London: Edward Elgar Publishing.
- Nalder, J. (2017). *Future-U A3 Model: how to understand the impact of tech on work, society and education*. Future-U. <https://www.nmc.org/blog/are-we-there-yet-artificial-intelligence-in-education>.
- Nilsson, N. J. (2010). *The quest for artificial intelligence —a history of ideas and achievements*. Cambridge: Cambridge University Press.
- Rosso, C. (2018). *The human bias in the AI machine*. New York: Psychology Today.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach* (3rd ed.). Upper Saddle River: Prentice Hall.
- Scott, R. (1995). *Institutions and organizations. Ideas, interests and identities*. London: Sage.
- Simonite, T. (2018). *Tech firms move to put ethical guard rails around AI*. San Francisco: WIRED.
- Stacey, R. (1992). *Managing the unknowable: the strategic boundaries between order and chaos*. London: Jossey-Bass.
- Still, R. R., Cundiff, E. W., & Govoni, N. A. (1958). *Sales management: decisions, policies, and cases*. Englewood Cliffs: Prentice-Hall.
- Tegmark, M. (2017). *Life 3.0*. London: Allen Lane.
- Thomsen, S. (2008). *An introduction to corporate governance: mechanisms and systems*. Copenhagen: Djof Publishing.
- Wooldridge, A., & Micklethwait, J. (2003). *The company: a short history of a revolutionary idea*. London: A Modern Library Chronicles Book.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Michael Hilb Is a Titular Professor at the University of Fribourg, Switzerland, and a member of various Boards of Directors and Trustees. He teaches strategy, entrepreneurship and corporate governance at universities in Asia and Europe.