Artificial Intelligence and Data Science Governance: Roles and Responsibilities at the C-Level and the Board

Bhavani Thuraisingham
Computer Science Dept.
The University of Texas at Dallas
Richardson, TX, USA
bxt043000@utdallas.edu

***Abstract*— Corporate governance and the roles and responsibilities of the corporate officers and the board of directors have received an increasing interest since the Enron scandal of the early 2000s. This scandal resulted in enacting policies, laws and regulations such as the Sarbanes-Oxley and others. More recently, with almost every corporation focusing on the applications of Artificial Intelligence (AI) and Data Science (DS) for their businesses in numerous industries including finance and banking, healthcare and medicine, manufacturing and retail and defense and intelligence, it is critical that these corporations take a serious look at the roles and responsibilities of the corporate officers and the board with respect to the governance of the AI and DS operations. This paper discusses the issues and challenges for AI and DS governance with an emphasis on the potential roles and responsibilities of the corporate officers and the board of directors.**

Keywords—Artificial Intelligence, AI, Data Science, Corporate Governance, Chief Artificial Intelligence Officer, Cyber Security

I. Introduction

Artificial Intelligence (AI) and Data Science (DS) are affecting every aspect of our lives from healthcare to finance to driving to managing the home. Sophisticated machine learning (ML) techniques with a focus on deep learning are being applied successfully to detect cancer, to make the best choices for investments, to determine the most suitable routes for driving as well as to efficiently manage the electricity in our homes. We expect AI to have even more influence as advances are made with technology as well as in learning, planning, reasoning and explainable systems [1]. While AI has become central practically to every industry, applications of DS are also exploding in every industry. Large amounts of data are being collected, stored, managed, processed and analyzed. For example, ML techniques (also referred to as data mining by many) are being applied to the data gathered so that intelligent decisions can be made. We believe that AI and DS intersect at ML [2]. That is, while DS includes areas such as gathering and organizing the data, AI includes areas such as planning and searching, ML is common to both of them and includes techniques such as neural networks, classifications and clustering.

While the advances in both AI and DS will greatly advance humanity, organizations such as the United Nations have embarked on initiatives such as “AI for Good”. In addition, there is also concern that AI and DS may be used for wrong purposes such as robots killing people without a reason and AI algorithms being biased or not being fair and resorting to discriminatory practices. This means that organizations must focus on how to use AI to enhance the business of the corporation without AI being biased or unfair or causing harm. That is, corporations must think seriously about the governance aspects of AI and DS. They need to also consider the risks involved in applying AI techniques as in many cases they produce incorrect solutions as well as result in false positives and negatives. This means that a corporation’s AI and DS strategies must be discussed at the highest level. What is a better way to ensure the governance of AI and DS than having a Chief AI Officer (CAIO) together with a Chief Data Officer (or both combined into one)? Such an officer should be a member of the corporate board and possibly report to the CIO (Chief Information Officer).

In a recent article, we explored the issues involved in cyber security governance and stressed the need and the importance of having a cyber security expert as a board member [3]. We argued that a business must be proactive about cyber security. Cyber security must have an advocate in the company. Ideally this should be the CISO (Chief Information Security Officer) reporting to the CEO. We also stressed that the cyber security strategies must be intertwined with the corporation’s business strategy. Based on our initial investigation of Cyber Security in the Boardroom, we are now focusing on the governance aspects of AI and DS. This paper first describes how AI and DS are being utilized by every industry to give them a competitive advantage. It then focuses on aspects of AI governance which will also include a discussion of DS governance. Next it discusses the integration of cyber security into the governance of AI and DS because every technology being deployed in a corporation must take into consideration its cyber security implications. Finally, it discusses the roles and responsibilities of various corporate officers and board members in promoting AI and DS.

The organization of the paper is as follows. First, we will briefly discuss the developments in AI and DS in Section 2. Next, we will discuss AI governance in Section 3. The impact of cyber security on AI governance will be discussed in Section 4. In

978-1-7281-1054-7/20/$31.00 ©2020 IEEE
DOI 10.1109/IRI49571.2020.00052

Section 5, we will discuss the roles and responsibilities of the corporate officers and board members to support AI governance. The paper is concluded in Section 6.

II. DEVELOPMENTS IN AI AND DS

Over the past decade the world has heard a lot about Data Science and especially Artificial Intelligence (AI). Whatever we read and watch, AI is mentioned in most texts and films. This section explores the developments in both AI and DS. Well before we had DS, we had Database Systems and AI since the 1950s. While the origins of AI go back to the seminal paper by Alan Turing in 1951 "Can Machines Think?" [4], early database systems in the late 1950s were based on the network model (e.g., Codicil). Both fields developed independently with expert systems and rule-based systems dominating AI and relational database systems dominating database systems in the 1970s [5]. Furthermore, ML was developing as a sub-area of AI at that time. In the meantime, efforts to integrate AI and Databases began in the late 1970s with logic databases and intelligent database systems where reasoning and inferencing were built into database systems. This integrated intelligent database management was a major area throughout the1980s and into the early 1990s. While database systems research continued to explode throughout the1980s and 1990s, AI was not progressing as much and did not meet expectations. For example, while MCC's Cyc project had high hopes, the objectives of the project were not met.

As the amount of data to be managed grew larger and larger especially with the WWW, the word data management was coined. It included not only database management but also managing the large amounts of data that included cleaning the data and ensuring the integrity of the data [6]. Soon in the early 1990s, the goal was to extract nuggets from the data so that intelligent decisions could be made. Statistical reasoning techniques (borrowed from Statistics) and machine learning techniques (borrowed from ML) were being applied to the data to extract useful patterns as well as predict new patterns. This field came to be known as *data mining*. Data mining grew leaps and bounds between 1995 and 2005 and provided corporations the ability to analyze the data and extract useful information. At the same time data warehouses that clean and aggregate the data to answer statistical queries became very popular. Data warehousing was considered by many to be the first step towards data mining. That is, one has to get the data ready to mine it and extract useful results [7].

However, in the early 2000s, DARPA's (Defense Advanced Research Projects Agency) TIA (Terrorism Information Awareness) project gave a lot of negative press for data mining as people were concerned their privacy was being violated [8] and so the United States Congress put a moratorium on data mining. Subsequently the name *data analytics* was used instead of data mining, Nevertheless the data mining area continued to thrive and flourish among the researchers.

In the 2000s the AI researchers were trying to advance machine learning by introducing many layers of learning in neural networks. This approach worked. Therefore, neural network research that was more or less dormant for some time soared with the concept of deep learning. By 2011 IBM released Watson and the popularity of ML was worldwide. Around that time, it was also possible to use cloud and high-performance computing to manage and analyze data [9]. Data management, data warehousing and data mining/analytics all merged into a field called *data science*. Data science included both getting the data ready for analytics as well as analyzing the data which included both statistical reasoning and machine learning. Essentially one could consider machine learning to be a part of data science. Data science is still very popular as without having the proper data you cannot apply the learning techniques effectively and produce accurate results.

Starting in the 2010s, artificial intelligence which was by then over 60 years old became hugely popular partly because the AI researchers always felt that ML was part of AI and ML had its origins in AI. After all, it was the AI researchers who invented deep learning in the 2000s. A major area of AI has always been ML. But AI also includes some less popular areas today such as expert systems and planning systems which were major disciplines of AI in the 1960s, 1970s and 1980s. AI also includes other areas such as multiagent systems and robotics. But it is the ML part of AI especially with deep learning that really took off in the 2010s [10].

Both AI and DS are playing major roles in the industry from healthcare to finance and manufacturing. The Data Scientists collect and clean the data as well as structure and organize the data for analysis. Various statistical reasoning techniques are being applied to this data to extract useful patterns. AI scientists then learn from the data and apply machine learning and especially deep learning techniques to make predictions about the future as well as to make the best possible decisions that also include detecting early stage cancers and preventing catastrophic events. While there has been debate as to whether ML belongs to DS or AI, it is now more or less accepted that ML is sort of at the intersection of AI and DS.

III. AI AND DS GOVERNANCE

Now that we have provided a brief overview of the developments in AI and DS, next we will discuss various aspects of AI and DS governance.

AI Governance: AI governance is the responsibility of the corporate executives and the board to ensure that the company follows all the processes and methods required for incorporating AI into their business operations including product development, manufacturing and sales. For example, is the company using AI techniques that have been tested? Do these techniques show bias? Are they fair? Do they harm the society? Are the AI Systems evaluated, certified and accredited? Do the developers and the users of the AI systems have the appropriate qualifications? These are questions that corporate leadership must address and obtain answers for. There are some efforts that have been reported on governance. Notable among them is Google's report on this subject. For example, Google has examined the following areas that are part of AI Governance that advance the legal and ethical aspects of AI [11]: 1. Explainability standards; 2. Fairness appraisal; 3. Safety consideration; 4. Human-AI collaboration; and 5. Liability frameworks. Additional areas we are investigating include *accountability* and *transparency*. Organizations must examine these areas and come up with an AI strategy that best suits their business needs and at the same time meets the fairness and other aspects discussed in the Google report.

Data Science Governance: In one of our earlier papers, we had discussed aspects of data governance [12]. In particular, we introduced the notion of data supply chain management and drew parallels between supply chain management and developing a data product. We also discussed information sharing in supply chain management and the risks and incentives involved for information sharing. We need to reexamine this work within the context of DS governance. That is, what is the relationship between DS governance and data governance? Is DS governance a combination of Data governance and AI governance? We believe that while data governance is an essential part of DS governance, DS governance also includes additional areas such as fairness and safety. As stated earlier, DS is about collecting, storing, structuring, managing, and analyzing the data so that useful patterns can be extracted. DS includes aspects of machine learning. Therefore, areas such as fairness in DS and trustworthy DS are being explored and are part of DS governance.

Protecting AI against cyber attacks and privacy violations: This should be a very high priority for every organization and should be considered as important as showing profits to the shareholders. Has the company developed appropriate protection and detection methods for the cyber attacks of the AI and DS techniques? Is it possible for the ML algorithms to analyze the vast amounts of data and subsequently violate the privacy of individuals? Are appropriate cyber security and privacy measures being adopted so that the AI techniques are trustworthy? Cyber security and privacy considerations are an important aspect of both AI and DS governance and will be addressed in Section 4.

Risks and insurance: Before developing AI techniques and incorporating them to the areas such as product development and manufacturing, the company has to carry out an in-depth risk analysis. If there are no risks that the AI techniques will be unfair and untrustworthy, then we have a perfect situation, However, in the real-world people have biases and there is a lot of unfairness and discrimination. These biases can be transferred into the AI techniques. Therefore, it is important for the corporation to carry out a thorough risk analysis and take out AI insurance so that the company is not sued for unfair practices resulting from the AI techniques. Based on the risk analysis carried out, the company can then develop an AI strategy that addresses areas such as. 1. Explainability standards; 2. Fairness appraisal; 3. Safety considerations; 4. Human-AI collaboration; 5. Liability frameworks; 6. Accountability; and 7. Transparency.

AI and DS strategy has to be integrated with the business strategy: Since AI and DS are being incorporated into the operations of almost every corporation, they have to be aligned with the business strategy. One cannot develop a business strategy that violates the legal and ethical considerations such as unsafe AI, AI with bias, and AI that is not fair. Similarly, one cannot focus entirely on fairness and bias at the expense of making the AI technique useless. That is, there has to be a balance between the usefulness of the AI techniques and its safety, fairness and bias. We need an AI expert who understands the issues involved surrounding policy and ethics and gives sensible advice to the corporation to develop an integrated business and AI strategy.

Extending the governance frameworks to include the board participation: Various frameworks have been developed for corporate governance. These include the framework developed by NIST (National Institute of Standards and Technology) as well as the COBIT (Control Objectives for Information and Related Technologies) framework developed by ICASA [13]. These frameworks specify the best practices for various cyber security management activities including cyber security governance. Such frameworks as well as the liability frameworks discussed in the Google report have to be examined for AI governance. Also, we need to include the roles and responsibilities of both the corporate officers and the board members in the development of the frameworks in addition to the roles and responsibilities of say the employees, the vendors and contractors. Such a framework will specify not only the best practices for AI risk analysis, insurance, as well as safety, but also what should the roles and responsibilities be of the officers such as the CEO, CFO, CIO , CISO, and the CAIO/CDO as well as the HR representatives and the board members including the chairman of the board and the designated AI expert board member. More details will be discussed in Section 5.

Evaluation, certification, accreditation and standards: One of the success stories of the cyber security field is the operation of evaluation, certification and accreditation of the secure systems. The first set of criteria for evaluating secure systems was published in the early1980s (called the Orange Book) [14] and now we have the Common Criteria. Once the systems are evaluated, then they are certified for use and then accredited by management. We need a similar approach for AI systems. That is, we need criteria for evaluating them and subsequently certifying them for use and then accrediting them. This way the systems can be evaluated for explainability, fairness, bias and other criteria. Also related is the development of standards for AI systems. In one of our earlier papers we proposed standards for data mining [15]. We need efforts in this area and NIST is in an excellent position to develop criteria as well as standards.

IV. SECURITY AND AI

Cyber security and AI/DS are two of the fastest growing fields in Computer Science and more recently they are being integrated for various applications [16]. They are being applied to practically every industry [17]. The collection, storage, manipulation, analysis and retention of massive amounts of data has resulted in serious security and privacy considerations. Various regulations are being proposed to handle big data so that the privacy of the individuals is not violated. For example, even if personally identifiable information is removed from the data, when data is combined with other data, an individual can be identified [18]. While collecting massive amounts of data causes security and privacy concerns, big data analytics applications in cyber security are exploding [19]. For example, an organization can outsource activities such as identity management, intrusion detection and malware analysis to the cloud. The question is, how can the developments in AI/DS techniques be used to solve security problems? Furthermore, how can we ensure that such techniques are secure and adapt to adversarial attacks? Our earlier work describes our research in data science including in stream data analytics and novel class detection and discusses its applications to insider threat detection [20].

While AI/DS techniques have many applications in cyber security problems, the cyber attacks can also occur on AI systems including machine learning and planning systems. Imagine the machine learning techniques being attacked and giving results that are erroneous. What happens if such systems are being used in pacemakers and automobiles? There is some work on an area called adversarial machine learning that adapts machine learning techniques to handle cyber attacks [21]. Furthermore, the machine learning techniques have the capability to analyze large quantities of data and extract patterns previously unknown. What happens if these patterns are highly sensitive information about individuals and subsequently violate their privacy? Data privacy should be of utmost importance to a corporation as data is the key to the success of a company. This is one of the main reasons that most corporations now have CIOs and some of them also have Chief Data Officers (CDO). These CIOs (and CDOs) must use the data as an asset so that the corporations generate profits. What is needed is a privacy-aware data lifecycle that uses the corporate policies for data collection, storage, management, analysis, sharing, retention and destruction [22].

V. RESPONSIBILITIES OF THE CORPORATE OFFICERS AND THE BOARD

In one of our earlier articles on cyber security governance, we stated that the business must have someone equivalent to a CISO. If a business is serious about protecting their data from cyber attacks, they must have expertise in cyber security. Board members do have legal and fiduciary responsibilities in the area of cyber security just like they would have in the area of finance. Cyber security strategy must be intertwined with the business strategy. The discussions in the previous sections have shown that like cyber security, AI and DS governance are also vital to the corporation. On the one hand the corporation must ensure that the sophisticated AI and DS techniques are applied to give them a competitive edge. This includes integrating AI and DS into areas such as medicine, finance, and manufacturing. Therefore, we need a CAIO to promote AI and DS in the corporation. Many companies now have Chief Data Officers. These CDO duties could be combined with those of the CAIO or they could be separate. What is also equally important is that the CAIOs must have a deep understanding of AI and the use of AI in an unbiased and fair manner. In addition, the CAIO must work with the CDO and the CISO to ensure that the AI techniques are not attacked or at least provide effective solutions to combat the attacks. The CAIO must work with the CFO and possibly the COO (Chief Operations Officer) to ensure that the AI strategies are intertwined with the business and financial strategies of the corporation.

While a CFO (Chief Financial Officer) is crucial for every corporation, the company must have a CAIO to work with the CFO to ensure that AI is used not only to benefit the company but it also must be used for the good of society. CAIOs have to be taken seriously and they must have authority and funding. Usually a CFO's job is to ensure that the company's finances are in order. Costs may impact the amount spent on AI solutions. This is why a company needs someone like a CAIO who is an advocate for AI. In the same way such an officer (or the CDO) must be an advocate for data science. Many corporations include cyber security as part of the responsibilities of the CIO (Chief Information Officer) or the CTO (Chief Technology Officer). We argued in our earlier paper that a corporation must have a CISO reporting to the CEO. In the same way we need a CIAO/CDO reporting to the CEO and be an advocate for AI and DS solutions and ensure that the techniques are fair and safe.

Finally, as I have mentioned, every corporation must have an AI and DS strategy. This strategy cannot be a standalone strategy. It has to be intertwined with the business and financial as well as the cyber security strategies. This is why the CEO, CFO, CIO, CISO, and the CAIO/CDO have to work together with the board members and every employee to ensure that the company provides appropriate AI solutions to improve its business but also ensure that the solutions are fair and unbiased. This means that the corporate board must have members who are knowledgeable about AI/DS so that they can ask the right questions during the board meetings.

VI. CONCLUSION

This paper has discussed the governance aspects of AI and DS and stressed the need and the importance of having an AI and DS expert as a board member. The question then is how does the company go about selecting such an expert? There are many people who now claim that they are AI experts. Therefore, the corporation has to make sure that the expert not only understands the technical issues of AI and DS, but also the policy, legal and ethical considerations. He or she has to understand not only the business aspects, but also needs to know the various ways the AI techniques could be attacked. That is, the expert has to work closely with the CISO of the company. He or she also needs to ensure that the AI techniques being employed by the company are fair, unbiased and trustworthy.

Finally, we would like to reiterate the key points made in this paper. A business must be proactive about the legal and ethical aspects of AI. AI must have an advocate in the company. Ideally this should be the CAIO (combined with the CDO). I believe that CAIO is as important as a CFO. They must work together. The board members have a legal and fiduciary duty to ensure that proper controls are being applied by the AI systems. The CEO and the business owners have the ultimate responsibility to hire the appropriate people and ensure that the CAIO gets sufficient funding to install appropriate solutions and ensure that the best practices in various areas such as 1. Explain-ability standards; 2. Fairness appraisal; 3. Safety considerations; 4. Human-AI collaboration; 5. Liability frameworks; 6. Accountability; and 7. Transparency are incorporated.

Acknowledgment

I thank my PhD students including Brin Ricks and colleagues Prof. Latifur Khan and Prof. Murat Kantarcioglu for their contributions to Security and AI. I also thank Ms. Rhonda Walls for editing this paper.

References

[1] Bhavani M. Thuraisingham: Can AI be for Good in the Midst of Cyber Attacks and Privacy Violations? A Position Paper. ACM CODASPY 2020.

[2] Bhavani Thuraisingham, what is the Relationship between AI and DS?

https://personal.utdallas.edu/~bxt043000/Motivational-Articles/Relationship%20between%20Data%20Science%20and%20Artificial%20Intelligence.pdf

[3] Bhavani Thuraisingham, Cyber Security and Data Governance Roles and Responsibilities at the C-Level and the Board, IEEE ISI, 2019

[4] Alan Turing, Can Machines Think, MIND 1950.

[5] Bhavani M. Thuraisingham, Data Management Systems Evolution and Interoperation, CRC Press 1997.

[6] Bhavani M. Thuraisingham, Web Data Management and Electronic Commerce, CRC Press 2000.

[7] Bhavani M. Thuraisingham, Data Mining: Technologies, Techniques, Tools and Trends, CRC Press 1998.

[8] Bhavani M. Thuraisingham*:* Data Mining, National Security, Privacy and Civil Liberties. SIGKDD Explorations 4(2)*, 2002*

[9] Bhavani M. Thuraisingham, Developing and Securing the Cloud, CRC Press, November 2013.

[10] Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, Pearson, 2009.

[11] AI Governance, Google Document, https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf

[12] Bavani Thuraisingham, "Data supply chain management: supply chain management for incentive and risk-based assured information sharing", *UT Dallas Technical Report*, pp. 36-11, 2011.

[13] Shon Harris et al., All-in-One CISSP, McGraw Hill, 2016.

[14] Department of Defense, Trusted Computer Systems Evaluation Criteria, 1983.

[15] Bhavani Thuraisingham and Chris Clifton, *Emerging Standards for Data Mining*, Computer Standards and Interface Journal (North Holland), Vol. 23, No. 3, 2001

[16] Bhavani Thuraisingham et al, Integrating Cyber Security and Data Science for Social Media: A Position Paper, IEEE IPDPSW, 2018

[17] Bhavani M. Thuraisingham, The Role of Artificial Intelligence and Cyber Security for Social Media, IEEE IPDPSW, 2020

[18] Bhavani N. Thuraisingham, Big Data Security, NITRD Workshop on Privacy, Washington DC, November 2015.

[19] M. Masud, L. Khan, and B. Thuraisingham, Data Mining Tools for Malware Detection, CRC Press, 2011.

[20] Bhavani Thuraisingham et al, Big Data Analytics with Applications in Insider Threat Detection, CRC Press, December 2017.

[21] Y. Zhou, M. Kantarcioglu, B. Thuraisingham, B. Xi, "Adversarial support vector machine learning", *Proceedings of the 18th ACM SIGKDD international conference on Knowledge Discovery and Data mining*, pp. 1059-1067, 2012.

[22] Bhavani Thuraisingham et al, Towards A Privacy-Aware Quantified Self Data Management Framework, ACM SACMAT Indianapolis, June 2018.