

DATA GOVERNANCE FIRMS IN CORPORATE BOARDROOMS

*Vanessa Villanueva Collao**

Project Summary

**WINIR Conference on Polycentric Governance and
the Challenges of the 21st Century**
Sept. 6–9, 2022

Abstract

Information/data-driven enterprises deal with a significant amount of information obtained from their users. These data need to be extracted, processed, categorized, and classified, which creates unmanageable datasets that quickly become outdated. Thus, data governance frameworks are crucial for management but especially for board decision-making. How to include data governance in the boardroom? This article illustrates how data governance can be implemented in the boardroom by creating a hybrid model: an “algorithm” next to the human board composition. This boardroom model includes a legal entity (a data governance firm) for designing a board algorithm consistent with the corporation’s purpose. A model situated at the intersection of a fully automatized and classical board composition.

INTRODUCTION

In June 2022, a computer scientist working on algorithmic bias at Google claimed that the chatbot, a program designed to emulate human conversations, had consciousness.¹ The news gained some notoriety after the employee published a transcript of his interview with the chatbot.² This

* JSD Candidate at the University of Illinois at Urbana-Champaign.

¹ Bobby Allyn, *The Google engineer who sees company’s AI as ‘sentient’ thinks a chatbox has a soul*, NPR:TECHNOL. (June 16, 2022, 4:31 PM), <https://www.npr.org/2022/06/16/1105552435/google-ai-sentient>.

² Blake Lemoine, *Is LaMDA Sentient?—an Interview*, MEDIUM (June 11, 2022), <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>

chatbot, known as LaMDA, short for Language Model for Dialogue applications, is a type of neural network architecture for language models. By mimicking biological neural networks, LaMDA can be trained to read many words, analyze their association, and predict what words will come next in the conversation.³ LaMDA is programmed to engage in a conversation receiving feedback from the interlocutor's responses.⁴

Once the technological underpinnings of LaMDA are deconstructed and there is an understanding of what the program does, Mr. Lemoine's observations raise some skepticism. Indeed, it does not seem that LaMDA was going through a malfunction but that the program (I oversimplify what a deep neural network is) properly achieved the purpose for which it was trained. LaMDA provided specific responses grounded on data collection.⁵

Data collection is part of our modern society. We constantly generate data. The data dimensions go from bare real-time information retrieval to daily interactions over social media to wealth creation using decentralized and permissionless systems that transform information into digital assets. Thus, the aggregation of data provides the input to re-conceptualize the corporate structure, facilitating data governance.⁶ At the same time, effective data governance requires management of large datasets and extensive training to take on those analytic techniques.

Today, the complexity of the board's tasks requires delivering informed decisions grounded on the systematic collection of data. The automation of data collection and analysis poses a threat to the classical human composition of a corporate board. In other words, is a board member replaceable by a non-human entity? However, the question should not be about 'board replacement.' Instead, the focus should be on how to successfully integrate a

³ Eli Collins & Zoubin Ghahramani, *LaMDA: our breakthrough conversation technology*, GOOGLE:AI (May 18, 2021), <https://blog.google/technology/ai/lamda/>.

⁴ LaMDA is pre-trained on 1.56 trillion words and 137 billion parameters. Given the great amount of information, this sophisticated statistical tool is able to predict accurate conversations because it possesses 40 times more words than previous dialogue models. *LaMDA: Towards Safe, Grounded, and High-Quality Dialog Models for Everything*, GOOGLE AI BLOG (Jan. 21, 2022), <https://ai.googleblog.com/2022/01/lamda-towards-safe-grounded-and-high.html>.

⁵ Google research showed that Transformer-based language models were able to be trained on dialogue and to interact with humans because they create dialogues by word association. However, these language models suffer from a common issue: the responses are not necessarily factual or correct. Collins & Ghahramani, *supra* note 3.

⁶ In this sense, data is characterized as the input of an intangible thing "that serves as the basis for a finding or decision." Jennifer Raso & Nofar Sheffi, *Data*, in THE ROUTLEDGE HANDBOOK OF LAW & SOCIETY 112 (2019).

useful analytical tool into the existing board structure.

The fear of technology reflects an overestimation of what technology can do and a lack of clear understanding of the relationship between humans and Artificial Intelligence (A.I.).⁷ Indeed, algorithms, deep neural networks, machine learning, and the like are all used interchangeably in common parlance and loosely grouped under the umbrella term of Artificial Intelligence (A.I.).⁸ However, there is a wide variety of techniques gravitating around A.I.⁹ and no unifying concept of what it embraces. Hence, the terminological choice of A.I. largely depends on what is sold (and to whom) rather than what it is. Most of the time, A.I. is used when referring to sophisticated statistical tools or programs.¹⁰

For corporations, sophisticated programs/software and their correct design can fulfill coordination and administrative tasks, which are crucial for management, efficiently solving problems, innovation, and engagement with stakeholders.¹¹ In addition, data collection can benefit the board because this

⁷ An interview report from FrameWorks showed people associating immediately the technology behind A.I. with robots and machines while its association with algorithms was less apparent. Lindsey Conklin, Emilie L'Hôte, Patrick O'Shea & Michelle Smirnova, *Communicating About the Social Implications of AI: A FrameWorks Strategic Brief* *Communicating About the Social Implications of AI 2 Communicating About the Social Implications of AI A FrameWorks Strategic Brief* 8 (2021), <https://www.frameworksinstitute.org/publication/communicating-about-the-social-implications-of-ai-a-frameworks-strategic-brief/>.

⁸ The term Artificial Intelligence was coined by the scientist John McCarthy during a summer workshop at Dartmouth college in 1955. John McCarthy, *COMPUTER HISTORY MUSEUM: PROFILE*, <https://computerhistory.org/profile/john-mccarthy/>. McCarthy defined A.I. as the science and engineering of intelligent machines or programs. John McCarthy, *What is AI?* (2007), <http://jmc.stanford.edu/articles/whatisai.html>.

⁹ A non-exhaustive list of the technologies incorporated into the umbrella term of A.I. are Automation, Biometrics, Decision Management, Deep Learning, Emotion Recognition, Image Recognition, Machine Learning, Machine Vision, Marketing Automation, Natural Language Processing (NLP), Robotics, Self-driving cars, Speech Recognition, Virtual Agents (or Assistants).

¹⁰ A recent post from Emily Tucker, the executive director of the Center on Privacy & Technology at Georgetown Law, highlighted the incorrect and misleading terminological uses of this “umbrella term,” and rejected its random evocation. Emily Tucker, *Artifice and Intelligence*, *MEDIUM* (Mar. 8, 2022), <https://medium.com/center-on-privacy-technology/artifice-and-intelligence%C2%B9-f00da128d3cd>. In this article, I will not employ artificial intelligence but directly refer to the specific technique or processes. For example, instead of machine learning I would refer to it as machine training which seems much more appropriate semantically and functionally.

¹¹ A survey of 1,770 managers from 14 countries revealed that on average 54% of their activities are administrative coordination and control. Vegard Kolbjørnsrud, Richard Amico & Robert J. Thomas, *How Artificial Intelligence Will Redefine Management*, *HARV. BUS.*

type of expertise is relevant to understanding trends and establishing projections of a future decline or shift in business models or markets—as happened when users migrated from one social media platform to another (from Facebook to Instagram to TikTok).

In general, corporations do not hire board services, and boards are formed by unaffiliated individuals. However, in information-driven enterprises hiring these services is becoming necessary because novel applications—and their interplay—fabricate larger and unmanageable amounts of data. Moreover, without an organized data framework, data analysis becomes irrelevant due to market conditions, regulation, incorrect classification, wrongful aggregation, etc. Likewise, enterprises are not accustomed to defining a hierarchy of importance and relevance of such data and how to develop a general approach to measurement adaptable across markets.

I argue that a non-human member, such as an algorithm,¹² should be incorporated as a board member for matters involving deep data analysis. To achieve this process, information-driven enterprises need the services and support of Data Governance Firms. These firms can assist in handling the algorithmic design and programming through involvement with management and other constituencies.¹³ Moreover, Data Governance Firms' expertise can assure consistency with the corporation's value but especially with the purpose of the corporation by introducing an algorithm that can propose ideas

REV. (Nov. 02, 2016), <https://hbr.org/2016/11/how-artificial-intelligence-will-redefine-management>.

¹² Computer science provides some definitions of algorithms as recursive models, defineable recursor [sic], or measurable functions or equations on a probability space, among other abstractions, without reaching consensus. Moshe Y. Vardi, Editor's letter, *What is an algorithm?*, 55 COMMUN. ACM 5 (2012). These definitions do little to inform and engage people in public discourse about its role. In an attempt to show what the algorithm is, I refer to a description used in the philosophy of computer science that describes an algorithm as a procedure that responds to a finite set of instructions and tells us what the next steps are to follow. Robin K. Hill, *What an Algorithm Is*, 29 PHILOS. TECHNOL. 35–59 (2016).

¹³ In the scholarly literature, there has been too much emphasis on shareholder primacy as the best way to pursue effective corporate governance, disregarding other constituencies. Recently, we have assisted in a new reading of corporate theory extending the corporate franchise to working-class shareholders and to stakeholders. DAVID H. WEBBER, *THE RISE OF THE WORKING-CLASS SHAREHOLDER: LABOR'S LAST BEST WEAPON* (2018). Thus, undermining contractarian theories supporting shareholder wealth as the main paradigm. The corporate franchise extension is not an abstraction but a consolidated reality in other systems whose experience can inform new ways to theorize about the firm. One such example arises from the comparative perspective of the German corporate experience. GRANT M. HAYDEN & MATTHEW T. BODIE, *RECONSTRUCTING THE CORPORATION. FROM SHAREHOLDER PRIMACY TO SHARED GOVERNANCE* (2021).

(namely, have a voice) and voting power.

Hiring an external firm, which designs data governance structures, also implements accountability. Data governance frameworks can take on *i.* monitoring costs (tracking performance), *ii.* bargaining costs (establishing frameworks where all the constituencies are included), and *iii.* coordination costs (enhancing coordination activity across constituencies).¹⁴ In addition, accountability relates to publicity. Open data can confer credibility and transparency to the algorithmic decision-making process.

It is not uncommon for corporations to depend on external firms or their internal units for advanced dataset analysis or machine training. However, when the results from data collection are delivered to the board, the board is entitled to discretion on whether to incorporate those results in their decision-making process, largely disregarding them. In this context, the Data Governance Firm would be in charge of observing, collecting, and storing information. These activities imply having a memory (a collective board memory)¹⁵ with the function of saving outputs both to compute (at the present moment) and how those outputs can be stored (past storage for future use). Thus, the Data Governance Firm will inform, communicate, and transmit the information providing interpretations to the board.

The incentive for boards to adopt this model is that Data Governance Firms can actively spot and avoid data biases. The issues with data bias are multiple and can relate to algorithmic bias (dealing with the algorithm's properties) or to data bias (involving the appropriateness of the data inputs).¹⁶ Much more popular in its connotation, data bias comprises three types of bias: sampling bias, prejudicial bias, and measurement bias.¹⁷

¹⁴ Innovations, like algorithmic tools, can reduce coordination costs. R. H. Coase, *The Nature of the Firm*, 4 *ECONOMICA* 386, 396 (1937). In information-driven enterprises, this reduction takes place by including different corporation constituencies in the data framework design, not merely aggregating them as a monolithic compound but distinguishing them and making their shades and nuances emerge.

¹⁵ M. TODD HENDERSON & STEVEN BAINBRIDGE, *OUTSOURCING THE BOARD: HOW BOARD SERVICE PROVIDERS CAN IMPROVE CORPORATE GOVERNANCE* (2018).

¹⁶ In describing the types of data biases, algorithmic bias relevance resides in the inadequacy of the model and the steps needed in order to achieve an optimal model that can accommodate complexities and is resistant to what statisticians call noise. Thus, limiting its ability to provide generalizations outside the training dataset. Jeffrey C. Chen, Edward A. Rubin & Gary J. Cornwall, *The Ethics of Data Science*, in *DATA SCIENCE FOR PUBLIC POLICY* 283–297 (2021). Contrarily, data bias regards whether there is good data or bad data in the model.

¹⁷ *Id.*

While sampling bias might become a problem leading to unsatisfactory or incomplete training data, prejudicial bias is a prime argument for the opponents of a data governance model because it touches on cultural or political issues regarding our understanding of the relationship between humans and machines. Precisely, prejudicial bias emerges when stereotypes influence data content or its classification.¹⁸ However, Data Governance Firms actively aim to train the model to spot these issues, halting a decision that can echo these prejudices.

Measurement bias, another type of data bias, deals with issues in setting up the parameters used to achieve a specific set of goals. Hence, measurement bias stems from the classification and coding of the information.¹⁹ A decision would likely be biased if the board simply takes algorithmic responses to base their choices without being involved in the algorithmic design or process.

Coding corporate rules opens up the discussion of defining the corporate purpose. A Data Governance Firm must be directly involved in design reliability, creating models that reflect measurable goals, and delivering consistent insights. Hence, in order to incorporate them, data-driven enterprises need to include explicit language in their charters to justify Data Governance Firms' privileges in developing an algorithm. An algorithmic board member is programmed to consistently take into account the corporate purpose, including goals that deal with service governance and the duties owed to beneficiaries and others.²⁰

Therefore, having a (non-human) member for machine training tasks is fundamental to the board's informed decisions because data analysis results can be connected to the corporation's purpose and goals.²¹ This purpose should be not only lawful or ethical (besides being profitable) but sustainable—especially for information-driven enterprises.²² Hence, it would

¹⁸ *Id.* at 289.

¹⁹ Eirini Ntoutsis et al., *Bias in data-driven artificial intelligence systems-An introductory survey*, 10 WIRE DATA MIN. KNOWL. DISCOV. 1–14 (2020).

²⁰ Elizabeth Pollman, *The history and revival of the corporate purpose clause*, 99 TEX. LAW REV. 1423–1452 (2021).

²¹ Inquiries on corporate purpose go back to the discussion between Dood and Berle regarding shareholder and stakeholder primacy. See E. Merrick Dodd, *For Whom Are Corporate Managers Trustees?*, 45 HARV. L. REV. 1145–1163 (1932); Adolf Berle, *For Whom Corporate Managers Are Trustees: A Note*, 45 HARV. L. REV. 1365–1372 (1932). This discussion around the role of the modern corporation remains vivid and revives attention to the function of corporate purpose clauses. Elizabeth Pollman, *The history and revival of the corporate purpose clause*, 99 TEX. L. REV. 1423–1452 (2021).

²² Sustainability and ESG matters are at the center of current debate and inquiry for corporations' action through a series of initiatives advancing regulatory change. Jill E. Fisch,

be possible to include working-class shareholders, satellite companies dependent on the firm, consumers, and other constituencies whose actions can be digitized in the board's decision-making process.

Including different constituencies in the board deliberation process is fundamental for future-thinking companies. For example, in the economics literature, information-driven enterprises are known as two-sided markets²³ because managing datasets for trade is more profitable than selling a product. In this sense, redefining what the corporation is—why it exists—and its role in society must stem from governance by design.²⁴ As a result, shareholders would not be the only constituency enforcing fiduciary duties. In other words, next to the duty of loyalty owed by the board to shareholders, in data-driven enterprises, the corporation (shareholders) also owes information fiduciary duties to platform users.²⁵ Eventually, by implementing data governance infrastructures, the corporation is fulfilling those fiduciary duties toward that constituency.

Once the purpose is defined, Data Governance Firms can provide a method to weigh all these factors whose results can inform strategy.²⁶ Therefore, this mechanism would eventually incentivize the development of a market for firms with expertise in machine training in combination with

George S. Georgiev, Donna M. Nag & Cynthia A. Williams, Comment Letter of Securities Law Scholars on the SEC's Authority to Pursue Climate-Related Disclosure (2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4129614 (commenting on the SEC proposed rules for "Enhancement and Standardization of Climate-Related Disclosures for Investors").

²³ These are markets where each agent (or side) interacts through a platform or intermediary, an information-driven enterprise. The role and structure of the platform are fundamental to increasing its volume of transactions and making a profit. Jean-Charles Rochet & Jean Tirole, *Two-sided markets a progress report*, 37 *RAND J. ECON.* 645–667 (2006).

²⁴ The corporate design must, in fact, stem from a "normative conception of corporate purpose which is [unrelated and] external to the technology." Christopher M. Bruner, *Distributed Ledger, Artificial Intelligence and the purpose of the corporation*, 79 *CAMB. LAW J.* 431–458 (2020).

²⁵ Andrew F. Tuch, *A General Defense of Information Fiduciaries*, 98 *WASHINGT. U. L. REV.* 1897 (2021).

²⁶ Research in this area is still evolving and growing. Until now, the mechanisms of the platform's development are board or management-oriented, which indicates that the perspective of trading personal data comes from secondary data owners. As a result, the main drawback is that this focus disregards primary data owners' perspectives (end-users). In this way, information-driven enterprises lose the opportunity to obtain feedback and build better structures that incentivize sharing personal data. Ahmed Saleh Bataneh, Rabeb Mizouni, Jamal Bentahar, May El Barachi, *Toward monetizing personal data: A two-sided market analysis*, 111 *FUTUR. GENER. COMPUT. SYST.* 435–459 (2020).

other tools.

At the same time, introducing an algorithm as a board member builds a precedent about what type of expertise an information-driven company's board needs to perform effectively. The algorithmic design is an intentional process because it is not enough to upload "good" data into the program and wait for a solution to appear magically. The result's validity does not solely depend on the quality of the data but on pre-established parameters for measurement. Thus, how the algorithm should be designed requires understanding the corporation's short- or long-term goals and the beneficiaries of that process.²⁷

There are multiple advantages of having a legal entity (such as the Data Governance Firm) helping the board. First, the model's productivity gains are reflected in the reduction of agency costs using machine training through data aggregation.²⁸ Second, the hybrid model provides concrete and accurate algorithmic responses to achieve fact-based decisions. Third, the Data Governance Firm can explain the data's measurement process and algorithmic design to the other board members. Most importantly, the Data Governance Firm's expertise can effectively monitor the risks of biased algorithmic programming and unsound results, not otherwise discoverable.

This hybrid model for boardrooms is situated at a middle ground between the full algorithmic and classical (human) board composition.²⁹ Accordingly, directors' tasks would not merely be relegated to increased oversight. Rather, the (human) director's decision-making process and analysis would be active in situations where the model does not fit, for instance, when data is missing,

²⁷ The scholarly debate has moved from promoting stakeholderism to a more holistic view of the interests involved in the corporation and cooperation from many fronts. ENACTING PURPOSE INITIATIVE, *Enacting Purpose Within The Modern Corporation. A Framework for Boards of Directors* (2020).

²⁸ John Armour & Horst Eidenmuller, *Self-Driving Corporations?*, 10 HARV. BUS. L. REV. 87 (2020).

²⁹ There are proposals to introduce automation on the board with a focus on replacement and entity/personhood attribution to the algorithm. Martin Petrin, *Corporate Management in the Age of AI*, 2019 COLUMBIA BUS. L. REV. 966–1030 (including algorithmic entities and fused boards/management); M. A. Tokmakov, *Artificial Intelligence in Corporate Governance*, in DIGITAL ECONOMY AND THE NEW LABOR MARKET: JOBS, COMPETENCES AND INNOVATIVE HR TECHNOLOGIES 667–674 (S. I. Ashmarina & V. V. Mantulenko eds., 2021)(including completely non-human controlled algorithmic board); Florian Möslein, *Robots in the boardroom: Artificial intelligence and corporate law*, in RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 649–669 (Woodrow Barfield & Ugo Pagallo eds., 2018)(theorizing about robots admission to the board replacing (human) directors).

creative thinking, and strategic matters.

The Data Governance Firm's role in developing an algorithmic seat is to transform the corporation and its constituencies into complex intersectional relationships based on data. Therefore, rather than being a *nexus of contracts*,³⁰ modern corporations and their constituencies can be included in datasets showing a *nexus of data* (in a better and harmonic way). Therefore, the proposal of an algorithmic component does not need to have greater weight than human components. Instead, it will help measure the amount of information the human and algorithmic seats have and give an ex-post account in light of whether the deliberation was sound and how the human seat decision differed from the board's decisions.

Indeed, this model allows the Data Governance Firm to be held accountable for the risks taken by the algorithmic seat while increasing transparency.³¹ Unlike independent directors, Data Governance Firms, as repeat players, have the incentives to monitor and oversee the algorithmic seat, creating reputational mechanisms and a robust market for corporate governance.³² As a result, this hybrid board model will provide a better management system addressing issues with the current shadow governance problems and loss of information.³³

Furthermore, this thought experiment also extends the debate on data.³⁴ Legal tools play an active role in deciding what is and what is not data. More critically, impending regulation can transform a working dataset into an outdated one. Nevertheless, data, as a whole, is still in the process of being regulated, and regulation can have a negative impact on data flows. For our

³⁰ Michael C. Jensen & William H. Meckling, *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, 3 J. FINANC. ECON. 305–360 (1976).

³¹ Consistent with the fiduciary duties the board owes to shareholders and those the corporation owes to users. Andrew F. Tuch, *A General Defense of Information Fiduciaries*, 98 WASHINGT. UNIV. L. REV. 1897 (2021).

³² The increasingly complex and risky operations of the board would incentivize board services for hire, adding expertise. STEPHEN M. BAINBRIDGE & M. TODD HENDERSON, *OUTSOURCING THE BOARD: HOW BOARD SERVICE PROVIDERS CAN IMPROVE CORPORATE GOVERNANCE* (2018). As some scholars have previously evidenced, new data analysis tools can provide meaningful assistance to the board. Sergio Alberto Gramitto Ricci, *Artificial Agents in Corporate Boardrooms*, 105 CORNELL L. REV. 869–908 (2020). However, the proposal of Data Governance Firms goes beyond mere assistance or technological support.

³³ Bainbridge & Henderson, *supra* note 32. In addition to competition to obtain an algorithmic seat on the board, these firms would provide precious information regarding corporate governance costs, rebalancing the information asymmetry between ownership and management.

³⁴ Some scholars have started to elaborate upon the social norms, definitions, and processes we as a society use when dealing with data. Raso & Sheffi, *supra* note 6.

purposes, and specifically for a modern corporation, data is a resource.³⁵ This resource is not goods or services³⁶ but social practices because people are constantly generating data.³⁷ Data are facts, and for this paper, corporate facts, as such facts are in the public domain and cannot be owned. However, there is a way to create data infrastructures with the specific purpose of promoting control and transparency through algorithmic tools.

This article is divided into four parts. The first part briefly overviews the different data techniques, how they are used for data extraction and analysis, and the importance of modeling and design. The second part explores the functions of the board, the Data Governance Firm's structure, and its connection with the corporate purpose. The third part examines how hiring data governance firms to develop an algorithmic board would be beneficial. The fourth part discusses how to implement coded corporate rules and the relevance of data governance frameworks to enhance the board. In the final part, the article concludes.

³⁵ Thomas Streinz, *International Economic Law's Regulation of Data as a Resource for the Artificial Intelligence Economy*, in *ARTIFICIAL INTELLIGENCE AND INTERNATIONAL ECONOMIC LAW. DISRUPTION, REGULATION, AND RECONFIGURATION*. 175–192 (Shin-yi Peng, Ching-Fu Lin, & Thomas Streinz eds., 2021).

³⁶ The most relevant agreement for goods and services is the General Agreement on Tariffs and Trade (GATT), which does not include data. GATT, Oct. 30, 1947, 61 Stat. pt. 5, 55 U.N.T.S. 194.

³⁷ Streinz, *supra* note 36.