JOLT DIGEST
HARVARD JOURNAL OF LAW & TECHNOLOGY

AI Comes to the Board Room in a Black Box: Are the Personal Assets of Directors at Risk in AI-Related Claims?

By Stuart Irvin, Seth Tucker, David Engvall & David Dapaah-Afriyie – Edited by Edwin Farley

Cite as: Stuart Irvin, Seth Tucker, David Engvall & David Dapaah-Afriyie, *AI Comes to the Board Room in a Black Box: Are the Personal Assets of Directors at Risk in AI-Related Claims?*, JOLT DIG. (Sept. 15, 2023), https://jolt.law.harvard.edu/digest/ai-comes-to-the-board-room-in-a-black-box.

Introduction

Artificial intelligence and machine learning ("AIML") technologies are transforming data-intensive industries, like healthcare and finance, at an astonishing speed. AIML tools can enhance decision-making processes at the enterprise level by analyzing financial data, operational data, customer data, and data collected in research and development activities.

The AIML tools and technologies that are beginning to come to market for enterprise customers can help management and boards to make more informed decisions based on real-time (or near real-time) insights. Business risks can be identified and mitigated, and business opportunities can be spotted and seized.

All new technologies, particularly when implemented at the enterprise level, carry a risk of loss for the business that implements them. Depending on the technology, the risks may include physical harm to the business's customers or even third parties, or they may be limited to financial losses incurred by the company. What happens when, for example, a pharmaceutical product that is developed using AIML tools results in poor health outcomes or patient deaths? What is the civil liability of the directors of a company if their AI-developed drug turns out to be the next thalidomide, or the outputs of AIML tools prompt the directors to pursue a business strategy that fails spectacularly? If history is any guide, the directors who rely on AIML tools could well face derivative actions brought by shareholders on behalf of the corporation alleging a breach of their duties to the corporation.

A focus of discovery in any future derivative action in this vein will likely encounter elements of the so-called "black box" problem. AIML technology, particularly in the context of deep learning models like neural networks, can develop so rapidly, and at such a level of complexity, that the internal workings of the AI model can no longer be understood by a corporation's management or even by the engineers who developed the model.[1] If the developers don't understand how an

[1] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889 (2018); Cynthia Rudin & Joanna Radin, *Why Are We Using Black Box*
(continued…)

AI tool is making decisions, it can be extremely difficult to correct errors or modify the tool to ensure its safe and ethical performance. At some level, the functioning of AIML technologies could, quite literally, be beyond human understanding.

To return to the "new thalidomide" hypothetical, would the directors of a pharmaceutical company have *personal* liability if the plaintiff in a derivative action can show that management and the board of the corporation knew, or should have known, that the drug development process was effectively a black box? In the business strategy hypothetical, would the directors face liability if their pursuit of a business strategy devised by a black box AI tool resulted in a plunge in the value of the shares of the company or even its bankruptcy? There certainly could be meritorious defenses to such claims, but the defense costs could be substantial and a judgment could be ruinous for a director who is named as a defendant in a personal capacity in the case.

Protecting Personal Assets

The personal assets of the directors who serve a corporation are typically protected by (1) indemnification commitments by the company and (2) Directors & Officers ("D&O") liability insurance. These protections will remain critical for individuals who serve as directors of corporations, including corporations that rely increasingly on AIML tools.

In this Commentary, we suggest updates to the standard indemnification terms used in contracts with directors and in corporate bylaw provisions. We also discuss best practices for ensuring that D&O policies continue to protect individual board members against the emerging risks that may follow from the widening use of AIML.

Information Systems and Red Flags

The use of AIML technology by a corporation to make decisions could expose the corporation's directors to two types of breach-of-fiduciary-duty claims. Directors serving on the boards of Delaware corporations owe fiduciary duties of care and loyalty, which include a duty of oversight. In the seminal *Caremark* decision,[2] the Delaware Supreme Court explained that the fiduciary duties of a director include a duty to make a good-faith effort to ensure that "information and reporting systems exist in the organization that are reasonably designed to provide to senior management and to the board itself timely, accurate information sufficient to allow management and the board, each within its scope, to reach informed judgments concerning both the corporation's compliance with law and its business performance."[3]

The Court also addressed when directors could be held liable for failing to implement a reporting system to facilitate board oversight. The Court noted that:[4]

> only a sustained or systematic failure of the board to exercise oversight—such as an utter failure to attempt to assure a reasonable information and reporting system exists—will establish the lack of good faith that is a necessary condition to liability.

Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition, HARV. DATA SCI. REV. (2019).

[2] *In re Caremark Int'l Inc. Derivative Litig.*, 698 A.2d 959 (Del. Ch. 1996).

[3] *Id*. at 970.

[4] *Id*. at 971.

> Such a test of liability—lack of good faith as evidenced by sustained or systematic failure of a director to exercise reasonable oversight—is quite high.

In *Stone v. Ritter*,[5] the Delaware Supreme Court ruled that to survive a motion to dismiss a failure-of-oversight claim, a plaintiff must allege particularized facts supporting a reasonable inference that either "(a) the directors utterly failed to implement any reporting or information system or controls; or (b) having implemented such a system or controls, consciously failed to monitor or oversee its operations thus disabling themselves from being informed of risks or problems requiring their attention."[6] A recent decision of the Delaware Court of Chancery has classified these two types of claims as, respectively, "Information-Systems Claims" and "Red-Flags Claims."[7]

The use of an AIML tool that functions as a black box for mission-critical decision-making tests the boundaries of both an Information Systems Claim and, as the risks associated with AI become better known to the public, a Red-Flags Claim. It is entirely possible that AIML tools could evolve to the point where the workings of a model, and its decisions that a corporation implements, can no longer be understood by a corporation's management or its board. On the basis of these facts, a court could conclude that the corporation has utterly failed to implement information and reporting systems that provide the board with timely, accurate information sufficient to allow the board to reach informed judgments concerning the corporation's compliance with law and its business performance. If whistleblowers, governance experts, or others publicize the risks to a corporation associated with these AIML tools, a Red-Flags Claim becomes more likely.[8]

---

[5] *Stone v. Ritter*, 911 A.2d 362 (Del. 2006).

[6] *Id*. at 370.

[7] *In re McDonald's Corp. Stockholder Derivative Litig.*, No.2021-0324, 2023 WL 387292, at *21–22 (Del. Ch. Jan. 26, 2023).

[8] As public understanding of a particular risk to companies increases, the potential for related red-flags claims increases—especially when the harm at issue falls directly within the ambit of a director or officer's responsibilities. *In re McDonald's Corp*. is illustrative of this dynamic. The case concerned a corporate officer responsible for ensuring workplace safety whom stockholder-plaintiffs alleged had consciously ignored red flags about sexual harassment and misconduct affecting company employees throughout his tenure with the corporation, which ended with the officer's termination for sexual harassment in 2019. *Id.* at. *2-5. In denying a motion to dismiss the stockholder-plaintiffs' red-flags claim, the Court highlighted widespread internal and external scrutiny of the corporation in connection with numerous allegations of sexual harassment, noting that this scrutiny contributed to a reasonable inference that the corporation had red flags for sexual harassment and misconduct of which the corporate officer was aware. *Id.* at *55-57. The increased public understanding of sexual harassment and the ability of directors and officers of a corporation to address and deter workplace sexual harassment factored into both the stockholder-plaintiff's complaint and the Court's decision. This greater public understanding, which in part enabled the red-flags claim in *In re McDonald's Corp*, is a relatively recent development, however, as the term "sexual harassment" was not coined until 1975 and did not enter the public consciousness until later. Sascha Cohen, *A Brief History of Sexual Harassment in America Before Anita Hill,* TIME (April 11, 2016). Moreover, it was only in 1986 that the Supreme Court ruled for the first time that a claim of "hostile environment" sexual harassment is a form of sex discrimination that is actionable under Title VII of the Civil Rights Act of 1964 (in *Meritor Savings Bank, FSB v. Vinson*, 477 U.S. 57, 63-69 (1986)). The history of red-flags claims related to sexual harassment suggests that as public understanding of the risk that uninformed reliance upon AIML (continued…)

Delaware law presumes that directors act in good faith, and to be viable a complaint must plead facts sufficient to support an inference of bad faith.[9] Establishing a breach of a duty of oversight "requires pleading and later proving disloyal conduct that takes the form of bad faith."[10] This is a high burden, but not an insurmountable one on the right facts.[11] And for directors, the prospect of personal liability means that even a small risk of a potentially catastrophic loss is highly problematic. It is for this reason that Delaware corporations typically indemnify their board members for the risks associated with derivative actions alleging a breach of fiduciary duties owed by directors to the corporation and purchase D&O insurance.

Indemnification for AI-Related Losses

Indemnification undertakings can be in the form of individual indemnification agreements or indemnification provisions contained in the bylaws or other charter documents of the corporation. Indemnification obligations also can be created by a vote of the board of a corporation or its shareholders.

These indemnification obligations are substantively quite similar, regardless of the form used to implement them, at least for companies chartered in Delaware. The corporation typically agrees to indemnify the director (called an "Indemnitee" in most provisions) if "the Indemnitee acted in good faith and in a manner that the Indemnitee reasonably believed to be in or not opposed to the best interests of the Company."[12] Some, but not all, indemnification provisions go further and define the concept of "good faith" in detail. A typical provision states:[13]

> Indemnitee shall be deemed to have acted in good faith if [Indemnitee's] action is based on the records or books of account of the [Company], including financial statements, or on information supplied to Indemnitee by the officers of the [Company] in the course of their duties, or on the advice of legal counsel for the [Company] or on information or records given or reports made to the [Company] by an independent certified public accountant or by an appraiser or other expert selected with reasonable care by the [Company]…In addition, the knowledge and/or actions, or failure to act, of any director, officer, agent or employee of the

---

tools by directors poses to companies increases, the potential for related red-flags claims will increase.

[9] *Id*. at 3.

[10] *Id*. at 2–3.

[11] For example, in *Marchand v. Barnhill*, reversing the Delaware Court of Chancery's dismissal, the Delaware Supreme Court permitted a failure-of-oversight claim to proceed against Blue Bell Creameries USA, Inc. and its directors. *Marchand v. Barnhill*, 212 A.3d 805 (Del. 2019). In permitting the claim, the Court held that the stockholder-plaintiff alleged sufficient facts to support a fair inference that the defendants failed to make a good-faith effort to establish a reasonable board-level monitoring-and-reporting system to ensure the exercise of due care with respect to an "essential and mission critical" compliance risk of the company: food safety. *Id.* at 824. In the absence of such a good-faith effort, the defendants would have breached of their duty of oversight. The Delaware courts never made a decision on the merits of the alleged breach, however, as the parties reached a $60 million settlement in July 2020 before trial.

[12] For a good example of an Indemnification Agreement, see the agreement filed by Gain Therapeutics, Inc. at the SEC, https://perma.cc/NZ5C-GU5C.

[13] *Id.* § 6(e).

> [Company] shall not be imputed to Indemnitee for purposes of determining the right to indemnification under this Agreement.

The question for directors is whether this indemnification language, which is intended to be very broad and to approach the limits of what is permissible for public policy reasons, is nevertheless sufficiently broad to cover decisions made by the corporation using AIML technologies, especially in circumstances where a black box problem is known or suspected.

An answer to this question would typically involve research under Delaware law and the careful crafting of arguments based on precedent that has absolutely nothing to do with AIML. While the common law can and does adapt to new and unanticipated circumstances, it also rarely provides clear "yes" or "no" answers, especially on issues of first impression. In addition, litigating a case to conclusion is expensive, and whatever is ultimately decided by a trial court will, for a time, be vulnerable to revision, correction, or reversal by subsequent courts looking at similar facts—a process that can extend for years.

Given the uncertainty inherent in the common-law process and the delay in getting guidance from the courts on issues involving a fast-moving technology, corporations may seek to avoid the problem entirely and expressly provide for good-faith reliance on AI-related decision-making. The exact standard to be used in indemnification provisions could be debated at length by corporate governance wonks and care has to be taken to stay within the bounds of what is permissible under state law[14] and public policy. But the process has to start somewhere, and to get that process started the authors offer the following addition to standard indemnification language and trust to the wisdom of the crowd to refine and improve upon it:

> Indemnitee shall be deemed to have acted in good faith if [Indemnitee's] action is based on the records or books of account of the [Company], including financial statements, or on information supplied to Indemnitee by the officers of the [Company] in the course of their duties **(including information that was created, in whole or in part, using deep learning, machine learning, or other artificial intelligence technologies)**….

The added text above is not intended to absolve a director from the duty of oversight. It merely seeks to confirm that a director can, in appropriate circumstances, rely in good faith on information created using AIML tools. The inclusion of language of this kind would help to counter the argument that reliance on information that was created using AIML technology is, in itself and without a further showing of dereliction, a breach of a director's duty to the corporation.

Ensuring D&O Insurance

Indemnification will usually be the first resort when a director is sued for an alleged breach of duty to the corporation. Indeed, D&O insurance often requires the corporation to indemnify directors against derivative claims to the extent permissible by law.

But if the corporation is unwilling or unable to indemnify its directors — as it might be if it were in financial distress — the individuals will look to the company's D&O insurance to cover their defense costs and any settlement or judgment.

[14] 8 Del. C. § 145.

At present, there is no standard, widely used exclusion that would bar coverage for a director accused in a derivative action of breach of a duty owned to the corporation arising from the corporation's reliance on AIML tools or the board's reliance on AIML outputs. The fact that the insurance industry has not developed AI exclusions is probably due to one or both of two reasons. First, the insurers that sell D&O policies understand that providing broad protection and keeping exclusions to a small number makes their product attractive, and that they will be at a disadvantage in the market if they lard their policies with exclusions. Second, AI and commercially available AI tools have burst into the public consciousness fairly recently, and the risks of AI implementation by corporations are only beginning to be understood. Insurers may not have yet had time to fully consider whether they wish to protect themselves by either excluding AI risk altogether or putting in place a lower limit (a "sublimit") for claims that arise out of an insured's use of AI.[15]

Finally, D&O policies typically have an exclusion for acts (and sometimes omissions) by a director that were deliberately fraudulent or deliberately criminal.[16] The safest way for directors to distance themselves from such an exclusion may well be to disclose to the public that a corporation is using AIML tools to automate certain business processes and that the members of the board are using AIML outputs to assist with decision-making. Such a disclosure could reduce the risk of a claim that a board member had acted "fraudulently" by accepting remuneration for board service but had outsourced all the work of being a board member to an AI tool. A disclosure of this sort might not prevent a breach-of-fiduciary-duty claim from being asserted, but it should mitigate the risk of the D&O insurer being able to avoid responsibility for the claim by invoking the exclusion for deliberately fraudulent conduct.

[15] Some insurance companies are moving faster to address AI-related risk of loss than others. Munich Re has advertised a policy called "aiSure™" that promises an "insurance-backed performance guarantee" for AI providers that "can increase your clients' trust in your AI solution." This is not a D&O policy, but as described it could provide insurance at a key link in the AI value chain, and the fact that Munich Re has brought it to market suggests that the industry could move quickly to implement new forms of coverage if there is a demand in the market for that coverage.

[16] Often, such exclusions apply only if a tribunal has determined in a non-appealable adjudication that the director engaged in deliberate fraud or criminal activity, or if the director has admitted such misconduct.