Fulfilling fiduciary duties in the AI era: emerging risks and responsibilities in AI-assisted corporate financial oversight

Jamilya PANABERGENOVA

Karakalpak State University

ABSTRACT

This article examines emerging legal issues and theories of liability for directors involved in the management of AI financial instruments that are protected as trade secrets. The main question of the article is whether excessive delegation of functions or lack of transparency of AI algorithms can undermine the performance of fiduciary duties by directors. By reviewing case law in the context of strict oversight of past technological failures, the article proposes a renewed approach to the use of blockchain tools that will maintain efficiency benefits while ensuring necessary reporting and accountability. The study suggests that governance based on the principles of auditing AI performance and setting minimum standards for explainability can help strike a balance between driving innovation, addressing liability issues, and aligning with modern doctrines that hold boards accountable for key decision-making. As algorithms become increasingly integrated into senior management decision-making processes, there is a need to further explore transparency mechanisms and monitoring processes that will support evolving fiduciary responsibilities about evolving automation capabilities that impact shareholder interests.

This is an open access article under the Attribution 4.0 International (CC BY 4.0) license (https://creativecommons.org/licenses/by/4.0/deed.ru)
Сунъий интеллект даврида ишончли вазифаларни бажариш: сунъий интеллект орқали корпоратив молиявий назоратда ўзага келадиган хавф ва мажбуриятлар

АННОТАЦИЯ

Хоҳирги кунда коорпоратив бошқарувда кузатув кенгаша аъзолари томонидан хўжалик жамиятларининг молиявий назорати, молиявий ҳисобот ва стратегик режалар тузиш жараёнинда сунъий интеллект тизимларидан фойдаланиш кўп учрамокда. Бироқ юоқин алгоритмлар муҳим функцияларни автоматлаштириганда ўзага келган юқори даражадаги носозликлар бошқарув хавфини ўзига олади. Ушбу мақола ривожланаётган ҳукмий масалалар ва тижорат сирлари билан қолланган сунъий интеллект молиявий воситаларини бошқариш билан ҳамкорият жараёнинда сунъий интеллекта юклав корпоратив бошқарув субъектларининг ўртасидаги фидуциар муносабатларнинг бузилишини аникқашдан иборат.

Натижалар шуни кўрсатадики, сунъий интеллектни текшириш бўйича принципларга асосланган кўрсатма, тушунтиришнинг минимал стандарти билан биргилик инновацион рагатланиши, жавобгарлик масалаларининг мувозанатлаштириши ва ғай технологиялар муҳим рол йўнайдиган қолдор учун жавобгар бўлган ғай таълимотларга жавоб бериши мумкин.

Алгоритмлар юқори менежментга чуқурроқ кириб борар экан, тадқиқотлар акциядорларга таъсир кўрсатадиган тез ривожланаётган автоматлаштириш имконийтлари билан боғлиқ ваколатларнинг ўзгарувчан мажбуриятларни бажарадиган шаффофлик механизмлари ва мониторинг жараёнлари янада аниклаш лозим бўлади.

Выполнение фидуциарных обязанностей в эпоху искусственного интеллекта: возникающие риски и ответственность в корпоративном финансовом надзоре с помощью искусственного интеллекта

АННОТАЦИЯ

В данной статье рассматриваются возникающие юридические вопросы и теории ответственности директоров, задействованных в управлении финансовыми инструментами ИИ, которые охраняются как коммерческая тайна. Основной вопрос статьи заключается в том, может
подотчетность совета директоров, корпоративное управление, управление рисками.

ли чрезмерная делегация функций или недостаточная прозрачность алгоритмов ИИ подорвать исполнение фидуциарных обязанностей директорами. Анализируя прецедентное право в контексте строгого надзора за прошлыми технологическими сбоями, статья предлагает обновленный подход к использованию инструментов блокчейна, который позволит сохранять преимущества эффективности при обеспечении необходимой отчетности и подотчетности. Результаты исследования показывают, что руководство, основанное на принципах аудита результатов работы ИИ и установление минимальных стандартов объяснительности, могут помочь достичь баланса между стимулированием инноваций, решением вопросов ответственности и соотвествием современным доктринам, возлагающим на советы директоров ответственность за принятие решений, в которых ключевую роль играют новые технологии. По мере того как алгоритмы все активнее интегрируются в процессы принятия решений на высшем уровне управления, необходимо дополнительно изучить механизмы обеспечения прозрачности и процессы мониторинга, которые будут способствовать выполнению развивающихся фидуциарных обязанностей в отношении эволюционирующих возможностей автоматизации, оказывающих влияние на интересы акционеров.

INTRODUCTION

Recent incidents involving automated financial systems have gone awry and highlighted new risks for corporate boards of directors in the AI era. In 2022, BlueBank incorporated a machine learning algorithm into its earnings projections and loan risk models with disastrous results – missing a major fraud scheme and materially overstating likely profits [1]. This led to SEC penalties, bank losses, and ultimately a shareholder lawsuit alleging BlueBank's directors breached their fiduciary oversight duties by blindly relying on flawed AI systems without sufficient governance. While AI tools can help boards analyze volumes of data, predict future performance, and maintain regulatory compliance, this promising technology poses both practical and legal challenges for upholding fundamental financial stewardship obligations.

This article explores the emerging risks and responsibilities of directors as AI becomes integrated into core financial control and reporting functions. Specifically, it addresses open questions around whether excessive delegation of judgment or inadequate transparency into automated systems may undermine directors’ duties under corporate law. It suggests that while properly governed AI could assist boards in fulfilling monitoring duties more effectively, new frameworks for transparency, testing, and maintaining human discretion are needed to align these tools with expectations around oversight capability and Caremark compliance programs [2]. A key thesis is that both counsel and courts should establish more precise expectations around when reliance on "black box" predictions or projections could lead to liability if flawed AI outputs mislead human judgment.
Background trends reflect rapid expansion of AI assisting or supplanting traditional financial roles–smart contract accounting systems, deep learning algorithms flagging suspicious transactions, and neural network earnings forecasts [3]. The core research problem is delineating new risks, requirements, and responsibilities for directors overseeing this technology given nondelegable fiduciary duties. Under what conditions could incomplete understanding or over-dependency enable faulty AI systems to effectively perpetrate or conceal financial misdeeds that boards should have discovered? In other words, how can Caremark obligations be adapted to reasonably govern automated intelligence liabilities?

To address these questions, this article examines parallel governance issues regarding expert systems and past technological disruptions. It suggests that while AI has unique risks, emerging case law on compliance systems offers insights into avoiding negligence through the deliberate balancing of machine and human capabilities. The article is structured as follows: Section 2 reviews literature on fiduciary duties, enterprise risk management expectations, and previous technological disruptions posing board oversight challenges. Section 3 summarizes the suite of AI financial tools increasingly adopted by public companies. Section 4 highlights transparency and reliance concerns created by AI systems. Section 5 puts forth liability theories if neural networks propagate or obscure financial misstatements or fraud. Section 6 concludes with recommendations for proper governance and proposes areas for further scholarship as algorithms penetrate deeper into the C-Suite.

**LITERATURE REVIEW**

Understanding the legal implications of AI financial systems first requires grounding in the established expectations around financial oversight and risk management duties. Seminal Delaware case law has delineated a corporate board’s responsibilities to implement compliance systems reflecting good faith efforts and sound judgment [4]. Similarly pertinent is scholarship dissecting parallels between emerging automated tools and past disruptive technologies that challenged governance norms.

**Fiduciary Duties and The Caremark Standard**
The foundation for shareholder claims alleging deficiencies in financial controls lies in fiduciary duties requiring directors' good faith attention to corporate affairs. As clarified in the Caremark ruling and its progeny, boards should make a good-faith effort to implement monitoring procedures, systems, and protocols reasonably aimed at keeping abreast of institutional risks [5]. This does not mean directors must have precise knowledge of all activities within the firm. However, especially for mission-critical functions like financial reporting, they must ensure "red flags" surface to the board through proper information and reporting systems such that they can exercise oversight judgments in good faith [6].

Caremark’s core standard is whether internal control systems reflect a board’s exercise of due care and compliance in good faith – if so, liability risk is minimal even if things go awry. But if internal systems are unreasonably lax or dysfunctionally designed, or if directors consciously ignore red flags, potential liability looms larger [7]. Much commentary has focused on properly adapting the “Caremark duty of loyalty” to modern challenges [8]. As algorithms assume expanding financial responsibilities, questions arise about whether flawed logic or excessive automation could enable circumstances violating the board’s oversight responsibilities.
Disruptive Technology and Management Theories. While AI-enabled financial tools create unique challenges, scholars have explored similar themes around board governance amidst past technological disruptions. Paralleling today's opacity concerns, research on the rise of complex expert and information systems (e.g. financial trading algorithms) in the 1980s examined whether overreliance or lack of comprehension inhibits directors' capability to exercise proper judgment as fiduciaries [9]. Courts have adopted expectations about compliance rigor and risk management systems as technologies and data storage evolve.

**AI SYSTEMS IN CORPORATE FINANCE**

As computing power and access to big datasets have grown exponentially, AI-powered finance tools have quickly proliferated across functions like accounting, reporting, projections, and more [10]. Public companies eagerly adopt these technologies to not just automate repetitive tasks, but to surface non-obvious insights hidden across massive, siloed systems and remarkably improve the accuracy of forecasts [11]. While AI holds promise to amplify human financial expertise, it introduces new opacity and dependency risks requiring governance adaptations.

**CORE AI APPLICATIONS IN CORPORATE FINANCE INCLUDE:**

Predictive analytics for financial planning – based on vast datasets, algorithms can detect subtle patterns and relationships invisible to human analysts. These insights are used to model likely cash flows, predict returns on large capital projects or investments, estimate distributions and ranges for future financial statement line items, and simulate the impact of strategic decisions. The main benefit is attempting to reduce uncertainty in budget planning and long-term projections [12]. However, if the training data fails to capture unlikely but impactful "black swan" developments, models will break down during economic shocks.

Anomaly detection for fraud prevention – AI profiling of normal transaction patterns, client behaviors, inventory flows, and other business data can automatically flag outliers, suspicious cases, or novel fraud vectors for investigation. By teasing out signals typically missed by rules-based systems, machine learning holds promise to massively expand fraud detection capabilities [13]. Yet the sophistication of AI models requires extensive data quality, engineering, and sandbox testing to avoid false alarms or new blind spots.

Unstructured data mining in financial filings – natural language processing now parses qualitative statements, linguistic signals and "soft" assertions buried within earnings calls, SEC disclosures and partner agreements searching for subtle early warnings of financial strains. Moving beyond numerical data analysis unlocks a trove of risk insights [14]. However, subjective language interpretation remains an enormous technical challenge.

Automated transaction classification and posting – "Smart contract" systems embed complex accounting rulesets to ingest invoices, inventory events, supply chain data and other structured financial activity. They automatically classify transactions and create appropriate journal entries to post [15]. This builds efficiency, but currently lacks flexibility needed for non-standard events.

Contract analytics – Much financial risk hinges upon complex terms within partnership agreements, insurance policies, loans and other contracts. Algorithms now assist in parsing these documents to model obligations, identify non-compliance, verify
satisfaction of performance milestones, and evaluate the degree of revenue certainty [16]. However considerable uncertainty still exists around comprehensive language understanding at a legal level.

In summary, while modern AI promises to amplify financial oversight, projections, and efficiency for boards and management through sophisticated data analysis, doubts emerge regarding the interpretability, robustness, and flexibility needed for directing firm strategy. This underscores why enhanced transparency and deliberate balancing of automated vs. human analysis is required.

**The opacity challenge of AI systems**

A persistent governance challenge posed by many advanced AI systems is an inherent lack of transparency in the logic driving their outputs, recommendations, and forecasts [17]. The sophisticated algorithms at the heart of machine learning tools are complex neural networks with thousands of parameters tuned through extensive iterative training on vast datasets. Their decision boundaries and reasoning cannot be reduced to simple human-interpretable "if-then" rules. This opacity risks inhibiting sufficient understanding for proper oversight [18].

Specifically, regarding mission-critical financial governance functions, if directors do not adequately comprehend what key relationships, patterns or signals an AI algorithm utilizes to generate earnings projections, risk thresholds, or flag unusual transactions, confidence in relying upon these black box tools diminishes substantially. Without reasonable visibility into critical model assumptions, real-world biases hidden in training data, weighting of key variables, etc., boards cannot fully assess inherent risks of distortions or failures. Alarmingly, researchers argue such algorithmic black boxes implicate serious emerging questions around legal liability and systemic harms from excluding human participatory decision-making [19].

Exacerbating transparency concerns are risks of complacent over-trust in or over-reliance upon AI systems due to their alluring sophistication, prior track record, or ability to uncover non-intuitive insights. Misplaced confidence replacing diligent monitoring and verification could slowly erode the capability of directors to detect distorted outputs or gradual performance degradation. Unrestrained automation threatens to inhibit prudent human judgment required to contextualize AI system outputs, evaluate nuances of unusual situations, question counterintuitive machine recommendations, and make sound interpretations before strategic decisions [20].

In summary, while AI promises to amplify human financial expertise, solely depending upon black box algorithms for core control functions without thoughtful constraints introduces dangerous gaps. Ongoing oversight mechanisms centered on human discretion are indispensable no matter how accurate or sophisticated AI systems may appear based on past performance. Further research should delineate how to strike the optimal balance between AI augmentation and preserving necessary human governance capabilities that maintain the legitimacy and comprehension of automated guidance [21]. New transparency methods and deliberate limitations on full automation are needed as algorithms permeate deeper into the C-Suite.

**Emerging theories of liability as AI risks propagate**

Given the central role of accounting integrity and financial controls in modern corporate governance, scenarios where AI systems propagate or allow new vectors for financial misstatements, misconduct or disruptive risk scenarios may increasingly spark
As case law continues to clarify expectations for directors amidst fast-moving technological disruption across once-stable institutional processes, several liability theories against fiduciaries are emerging if flawed AI contributes to reporting errors, projecting failures, or unprecedented fraud losses.

Negligent Oversight. Shareholders could allege that boards neglected ongoing oversight responsibilities to understand the capabilities and embedded limitations of integrated AI systems guiding critical forecasting, fraud prediction, or auditing analysis. Ignoring transparency requirements, suitability standards, or verification mechanisms for complex black box technologies directly inputting financial statements or strategic plans could demonstrate negligence in safeguarding data quality and interpretation. Without reasonable monitoring protocols to watch for distortion risks unique to machine learning, directors may struggle to show good faith governance efforts per Caremark standards [22].

Reckless Misrepresentation. In scenarios where earnings estimates, fraud signals, projections of capital returns, or other shareholder communications substantively informed by non-transparent AI systems prove inaccurate, unreliable, or deeply misleading following some market shock or data shift, investors may claim directors had no reasonable basis to endorse and disseminate unreliable system outputs lacking explainability safeguards or identified limitations. Positioning complex black box technologies as comprehensively accurate or dependable when their training constraints suggest substantial hidden risks could be viewed as reckless communication, putting directors further from business judgment protections [23].

Inadequate Risk Monitoring. Separately, if deployed AI models exhibit previously unknown training gaps only visible following market shifts, causing compliance breakdowns or introducing new monitoring blind spots that allow risk scenarios such as fraud or collateral damage from misconduct to spread rapidly without detection, Caremark precedent around board responsibility for risk information systems may support viable negligence claims. If internal controls were antiquated or unreasonable for new threats propagated by algorithmic tools depended upon by management, fiduciary duties around oversight modernization may establish litigation pathways absent efforts to keep monitoring capabilities aligned with cutting-edge dependencies [24].

In conclusion, as algorithms powered by vast data become deeply integrated across financial reporting infrastructure, courts will confront novel questions regarding the scope and continued evolution of monitoring duties in fast-changing technological environments. But substantial guidance exists in earlier cases governing board oversight modernization needs, internal control upgrades for new systems dependencies, and business judgment deference balanced with good faith requirements to reasonably inform human oversight capabilities amidst disruption. The growing integration complexity posed by AI warrants ongoing legal and governance scholarship to delineate updated expectations.

CONCLUSIONS AND RECOMMENDATIONS

As this article has explored, AI financial systems create enormous opportunities for efficiency gains, informed projections, and risk insights – but also novel governance challenges around ethical data usage, output interpretability, and evolving oversight obligations. Based on the preceding liability analysis, several conclusions and recommendations emerge:
Verification Processes. Boards should require ongoing verification processes for AI outputs affecting financial statements or shareholder communications before reliance or disclosure. Internal auditors must evaluate model logic, assumptions, training data, and monitoring protocols.

Explainability Standards. Directors should establish minimum explainability standards governing financial AI procurement, customization, and deployment. Vendors should provide details on major model drivers, limitations, and potential failure modes understandable to fiduciary stewards lacking advanced technical skills.

Staff Proficiency Requirements: Board and senior managerial training must cover AI topics – transparency needs, dependence pitfalls, early performance deterioration signals, and oversight methods to ensure human governance maintains effectiveness over algorithms.

Oversight Technology. Solution architects should provide dashboards visualizing performance metrics, data drift, and other trends enabling directors to monitor for degradation and adapt policies accordingly. This "meta-data" layer supports Caremark duties in the algorithm age.

Liability Clarity. Courts should continue clarifying acceptable reliance standards for board oversight of AI systems. Given inherent opacity risks, guidance is needed on governance mechanisms, training expectations, and monitoring vigilance Required to enable business judgment rule protections.

In summary, realizing the benefits of AI in finance requires adapting both internal policies and legal doctrine to address risks posed by opacity and over-dependence. Further research should delineate how fiduciaries can discharge evolving Caremark duties as algorithms permeate deeper into the institutional nerve center guiding strategy.

REFERENCES:
5. Guttman v. Huang, 823 A.2d 492, Court of Chancery of Delaware, 5 May 2003.


