

Fiduciary Black Boxes: Corporate Veil Doctrine in the Age of Agentic Decision Systems

Siddharth Kanojia

Middlesex University Mauritius, Flic en Flac, Mauritius

s.kanojia@mdx.ac.mu

<https://orcid.org/0000-0002-1479-5292>

Abstract

While corporate law has traditionally allowed a certain level of delegation, it has never had to address the problem of how to respond when the delegate isn't a human being. Contemporary agentic artificial intelligence systems conduct negotiations, perform pre-employment screening, execute transactions, and distribute investment capital with very little human involvement in processes that are, by their very nature, human processes and fiduciary processes. If the system acts in the context of ordinary business activities, but violates the duties of care and loyalty owed by a human officer, there is no coherent legal theory for holding anyone accountable for the violation. This is not a purely theoretical problem for the future; discriminatory algorithms in recruitment, unauthorized use of algorithms in trading, and black box investment allocation mechanisms have already produced tangible harms that cannot be attributed to any human agent under current law. The thesis of this paper is that the corporate veil and the fiduciary framework that surrounds it provide the proper locus for addressing this problem, provided that courts and lawmakers are prepared to rethink the duties of director oversight rather than merely falling back on formalism or AI personhood.

The consequences of this process going awry would be just as legal as they are theoretical, since standards of corporate veil piercing and fiduciary responsibility vary significantly between common law jurisdictions, and the use of agentic AI systems advances most rapidly in those same jurisdictions: the US, UK, Singapore, and increasingly India – which have developed their doctrine of algorithmic accountability least extensively. The approach taken by this paper, then, is purposefully constructed as an analytical framework to travel internationally rather than a jurisdictional reform proposal; indeed, if nothing else, it hopes to contribute to a comparative vocabulary for agentic liability.

Methodology and Structure

This paper develops by way of doctrinal analysis of key cases dealing with corporate veil, and fiduciary duties, alongside the emerging body of literature regarding technical and regulation of agency and black box AI. This paper will not undertake an empirical investigation of the harms that algorithms may cause; rather, it will examine some specific examples of discriminatory hiring algorithms, and unauthorized algorithmic trading, as tests of the sufficiency of respondeat superior, AI personhood, and the proposed hybrid model. This paper is structured in relation to three questions: How does the existing doctrine define liability in cases in which the wrongdoer is not a person but rather a non-human entity? Why do the two major reform suggestions, put forward by the literature, not resolve this liability gap? And what would be the minimum doctrinal modification to close the liability gap?

The Doctrinal Gap

In the classic liability tool kit, the underlying assumption is that there is a responsible mind somewhere in the chain of causation. The doctrine of respondeat superior, for example, makes the employer vicariously liable for his employees' tortious conduct because the employee is an intentional actor, a negligent one, or an actor who at least acts with an identifiable state of mind that is open to legal characterisation. Similarly, fiduciary breach is premised on the capacity of the trustee to exercise judgment and balance competing interests. The agentic decision system, however, upsets this structure in two ways at once. For one thing, these systems simulate a human being's functional capacity as an employee or officer closely enough that injured parties would look instinctively to corporate law for help but without any mens rea, thus the legal foothold for respondeat superior disappears. For another, the agnostic nature of many such agentic systems, especially those based on machine learning architectures, makes the chain of causation between the relevant design/implementation decision and a downstream harm hard to reconstruct. The result is an accountability void: harm occurs, a human beneficiary of the corporate structure profits from the system's operation, but no individual satisfies the causal and mental-state prerequisites that existing liability rules require.

Why the Available Alternatives Fail

There are two prevailing approaches in the existing literature, and this paper posits that both are flawed. The first is the creation of a form of personhood or agency for the artificial intelligence itself, which is advanced as a solution that would provide a defendant for the courts to pursue litigation against. In this paper, it is argued that personhood solutions actually achieve the exact opposite of their intended effect: by creating an empty entity without any means of holding assets, continuity of intention, or even deterrence, personhood becomes a liability sink, further distancing the human actors responsible for authorizing the system's use and able to prevent recurrence.

The other is to let respondeat superior and standard fiduciary review alone and rely on the flexibility of negligence law to evolve on a case-by-case basis. This article contends that such an approach misapprehends the evidence imbalance inherent in black-box opacity: when a plaintiff or minority shareholder contests a board's oversight of a black-box algorithmic system, she confronts a discovery process that standard business judgment deference could not possibly have anticipated, and which grants immunity from liability to boards that employ such systems.

A Hybrid Liability Framework

The paper's central contribution is a three-part hybrid framework designed to sit within, rather than replace, existing corporate law architecture.

First, the implementation of a capped indemnity arrangement would enable organizations to set out the level of financial exposure involved in the deployment of such systems *ex ante*, akin to director and officer insurance, but tailored specifically to the risk inherent in the use of algorithms rather than simply the error on the part of managers. This ensures that the company still has an incentive to automate, but at least some of the cost is internalized.

The second recommendation is the introduction of a rebuttable presumption of the negligence of the directors in case the agency system governed by the board leads to a breach of the fiduciary duty that would normally result in the liability of a human officer. Instead of making the plaintiffs prove the causation using the difficult-to-understand model, the burden of proof lies on the shoulders of the directors to show that proper systems of governance, technical control, escalation procedure and risk analysis were established.

Third, this proposal mandates the performance of algorithmic audits on a mandatory basis and at the board level, where they must be recorded in the board's fiduciary record, in the same way that financial audit committees perform their functions. Such audits do not necessarily demand that the board members have any understanding of the inner workings of the algorithm, but do demand that they engage substantively with the risks and limits of the algorithm if they wish to avail themselves of the business judgment rule.

Preserving Salomon While Expanding Oversight

One consistent concern with the development of this body of literature is that extending exposure of directors for any AI-generated wrongdoing runs the risk of undermining limited liability altogether. In this paper, however, we hold the exact opposite view, and the hybrid structure has been deliberately crafted to preserve the separation of legal personality as articulated in the *Salomon v Salomon* decision. Limited liability shields the shareholders from the liabilities of the company, and it was not intended to shield directors from their obligations in respect of the use of the technologies they decide to use for purposes of the company. Through placing the new obligation at the level of the obligation to oversee rather than the level of the corporation itself, the hybrid model achieves accountability for the agentic system without undermining the separation of the company and the shareholder, which lies at the heart of the modern conception of the corporate entity.

Contribution to the Conference Theme

The paper directly addresses the liability and governance challenges facing the conference through its doctrinally conservative yet substantially impactful proposed change: one that does not ask legislatures to construct new classes of legal persons and one that does not ask courts to stretch existing tort law or fiduciary standards beyond their evidentiary boundaries. The following paper presents an approachable analytical framework, limited exposure, shifted burden of proof, and required audit at the board level that can be applied across common law jurisdictions regardless of how far behind they are on corporate governance and AI legislation. This paper also raises the question of how the standard for negligence needs to be set at the point where agentic systems reach full autonomy.

More generally, this paper frames corporate fiduciary doctrine as a tool too infrequently utilized in the broader discussion of governance and liability that is part of the agenda for this conference. In many ways, the bulk of the extant agentic artificial intelligence liability literature is taken from either tort law or sector-specific AI liability regimes, both of which have the risk of having to build entirely new regulatory frameworks from scratch. Corporate law is already well-versed in the language needed to assign responsibility between a principal and the instruments it uses to conduct itself through them. The proposed hybrid regime implies that the more sensible course of action would be to adapt such language, rather than replace it altogether.

Keywords

Corporate Veil; AI Agents; Fiduciary Duties; Black Box Algorithms; Director Liability; Limited Liability; Algorithmic Accountability; Digital Employees

Bibliography

- Tsang, K. F., & Fan, K. (2023). The proper role of agency in corporate group. *Int'l Law.*, 56, 333.
- Kring, B., Lee, S., Curtis, S. K., Stalley, L., Ventura, T., Raka, N., ... & Ahsun, A. (2025). Bridging the Trust Gap Between Black-Box Predictive Algorithms and Human Fiduciary Compliance.
- Harper, S. B., & Weber, E. S. (2022). Fiduciary Responsibility: Facilitating Public Trust in Automated Decision Making. *Journal of Social Computing*, 3(4), 345-362.
- Gangavarapu, R. (2025). AI governance: preparing for the rise of Agentic AI. In *Mastering AI Governance: A Guide to Building Trustworthy and Transparent AI Systems* (pp. 111-119). Cham: Springer Nature Switzerland.
- Henderson, M. D. (2025). Agentic AI and the ethics of leadership maintenance: rethinking responsibility in algorithmic organizations. *Leadership & Organization Development Journal*, 1-15.