

THE DISTINCTION

The Two Duties: Care and Loyalty in the Age of AI

Boards have carried the duties of care and loyalty for a century. AI does not add a third duty. It changes what each of the two requires, and what the business judgment rule will protect.

Brian J. Allen · Executive Director, Center for AI Oversight · August 2026

THE BOTTOM LINE

The duty of care asks whether the board did the work to be informed. With AI, that work is discharged through an oversight program, and the questions never asked are where liability lives.

The duty of loyalty asks whose interests a decision served. With AI, a conflict no longer needs a conflicted person. It can live in the configuration of the system itself.

The business judgment rule protects judgment that was actually exercised. Judgment surrendered to a system nobody interrogated may earn no protection at all.

Every director knows the two duties. The duty of care requires diligence: inform yourself before you decide. The duty of loyalty requires fidelity: decide for the corporation, not for yourself or anyone else. For a century these have been the twin pillars of fiduciary law, and most directors could recite the distinction from memory.

AI tests each pillar in a different way, and the difference is worth understanding precisely. In a recent article, *Fiduciary Duties and Business Judgment Rule 2.0 in the AI Act Age*, corporate law scholars Maria Lillà Montagnani and Maria Lucia Passador argue that AI regulation, led by the EU AI Act, is quietly reshaping both duties and the judicial deference that sits on top of them. Their analysis is European in origin and academic in register, but the logic travels well. What follows is that logic, translated for the boardroom.

Care: Did the Board Do the Work?

The duty of care has never demanded that directors be right. It demands that they be diligent: that decisions rest on reasonable information, reasonable inquiry, and reasonable attention. Delaware extended this to oversight in *Caremark* and sharpened it in *Marchand*: a board must have working systems for surfacing mission-critical risk, and failing to build them is itself the breach. No bad decision is required. The absence of the structure is enough.

Applied to AI, Montagnani and Passador call the result a duty of AI due care, and their most useful contribution is a plain taxonomy of how boards fail it. Procedural neglect: the risk structures, documentation, and monitoring were never put in place. Epistemic abdication: the structures exist, but nobody interrogates the system. The vendor's assurances are accepted, the model's limits go unexamined, the questions are never

asked. Structural complicity: the organization deploys anyway, accepting opacity or bias as the price of speed.

Notice what connects all three: omission. The care failures that will define AI litigation are not reckless decisions. They are questions that were never asked, escalations that never happened, records that were never kept. As the authors put it, silence in the face of technological opacity is not neutrality. It is exposure.

None of this requires directors to read code. The standard emerging is closer to what the authors call cognitive adequacy: knowing which questions matter. What can this system not do? Where did its data come from? What happens when it is wrong, and who finds out? A board does not discharge this duty personally. It discharges it through a program that asks those questions continuously and keeps the evidence. That is the Informed Decision Standard at work: courts will not demand the right answer, but they will demand proof the institution was positioned to make an informed, defensible decision.

Loyalty: Whose Interests Did the System Serve?

Loyalty asks a different question entirely. Not whether the decision was informed, but whom it served. The classic breaches are familiar: self-dealing, undisclosed conflicts, the director on both sides of a transaction. Boards have well-worn machinery for this: disclosure, recusal, independent committees.

AI relocates the conflict. The machinery watches the people in the room. It does not watch the systems the decision passed through on its way there. A vendor's model may encode optimization preferences that quietly favor

the vendor. A recommendation engine may be tuned, deliberately or not, toward outcomes aligned with an interest no one disclosed. A system built to maximize a single metric will silently filter out the considerations directors are legally bound to weigh, and the output will arrive at the board looking like neutral analysis.

This is what Montagnani and Passador call AI loyalty oversight: an affirmative obligation to ensure that the systems through which corporate judgment is exercised are themselves unconflicted and aligned with the corporation's interests. The unsettling part, and the reason it deserves board attention, is that loyalty can now be compromised without anyone intending disloyalty. The law has long held that a breach does not require bad intent; the old English fiduciary cases imposed liability on directors who acted within a conflicted terrain even while believing they served the company. Algorithmic delegation makes that terrain far harder to see.

Care asks whether the board understood the decision. Loyalty asks whom the decision served. AI can defeat both without anyone acting in bad faith.

The distinction, then, in one line each. Care is about the quality of the decision process. Loyalty is about the integrity of the interests it serves. And a board can fail one while satisfying the other: a flawlessly diligent process running on a conflicted system breaches loyalty; a perfectly aligned system nobody ever questioned breaches care.

The Business Judgment Rule: A Shield That Is Earned

Why does the distinction matter now? Because of what sits on top of both duties: the business judgment rule. Courts will not second-guess a board's honest, informed decision, even one that turns out badly. That deference is the space in which boards take risk, and it rests on a single premise: that judgment was actually exercised.

Delegating a decision to a system the board never vetted breaks that premise. Courts have already withdrawn deference from boards that relied blindly on conflicted financial advisors; Montagnani and Passador argue, persuasively, that unexamined AI will be treated the same way. Their term is BJR 2.0: the shield survives, but it must be earned in advance, through what they call demonstrable stewardship. Escalation paths for algorithmic anomalies. Records of what was asked and what was known. Evidence that when the system recommended one thing and judgment said another, the disagreement was noticed and resolved by a person with authority.

Read that list again and it describes an oversight program. Which is the point the Center has argued from the start: the program is the governance. The same architecture that discharges the duty of care and protects the duty of loyalty is the architecture that preserves the board's protection under the business judgment rule. This is not defensive lawyering. The institutions that can demonstrate their judgment are the ones that get to keep exercising it, at speed, while their competitors rediscover caution the hard way.

What This Means For Your Board

None of this asks the board to evaluate models or approve algorithms. Both duties are discharged through the program management runs and the evidence it produces. These are questions to put to management.

-
- If a decision informed by AI were challenged tomorrow, could management produce the record: what was asked, what was known, and who decided?
-
- When management relies on a vendor's system, who interrogated its limits, its data, and its objectives, and what did they find?
-
- Who sets the objectives our AI systems optimize for, and how would we learn if those objectives drifted from the corporation's interests?
-
- Do the systems supporting consequential decisions weigh what the law requires the corporation to weigh, or only what is easiest to measure?
-
- Where judgment departed from an AI recommendation, is that departure on the record? Judgment that leaves no trace earns no deference.

This essay draws on Maria Lillà Montagnani and Maria Lucia Passador, “Fiduciary Duties and Business Judgment Rule 2.0 in the AI Act Age,” which develops the concepts of AI due care, AI loyalty oversight, and the taxonomy of governance failures discussed above in light of the EU Artificial Intelligence Act (Regulation (EU) 2024/1689). Judicial framing draws on *In re Caremark International Inc. Derivative Litigation* (Del. Ch. 1996), *Marchand v. Barnhill* (Del. 2019), *In re Del Monte Foods Co. Shareholders Litigation* (Del. Ch. 2011), and *In re Rural/Metro Corp. Stockholders Litigation* (Del. Ch. 2014).