Integrating AI into Strategic Business Decisions: A Risk Management Perspective

Runlong Zou

The University of New South Wales, Sydney, New South Wales, Australia
zourunlong27@163.com

Abstract. Companies have started incorporating AI into decisions concerning acquisitions, market entry, and resource allocation, yet less attention has been paid to how deeper AI involvement changes the firm's risk profile. This conceptual study adopts an enterprise risk management perspective and draws on prior experimental, organizational, and regulatory evidence, including published research on reliance on algorithmic recommendations and an IOSCO regulatory survey that received 184 responses. It develops a framework distinguishing technical, decisional, and institutional exposure and relates these layers to the organizational locus of responsibility. AI involvement is represented by three analytical anchors (informing, recommending, and deciding) on a continuum of decision authority. The framework proposes that absolute technical exposure increases and then approaches a control-dependent ceiling, that decisional exposure is greatest when AI recommends while humans retain formal authority, and institutional exposure increases as AI outputs become more directly connected to organizational action. These trajectories are theoretically expected patterns evaluated against prior evidence rather than estimates derived from newly collected data. The resulting calibration principle links permissible AI authority to decision reversibility, meaningful human override, and accountable ownership.

Keywords: Artificial Intelligence, Strategic Decision-making, Enterprise Risk Management, Accountability, Corporate Governance.

1 Introduction

Acquisition screenings, market entry prioritization, and capital allocations no longer go directly to those responsible for them without first passing through models, and every such step means that some risk from the firm goes somewhere beyond the scope of control set up for the function at hand. The transformation has been witnessed more than speculated about, as firms increasingly integrate AI into strategic decision-making processes. That hesitation is generally attributed to thin skills, unclear cost, and uneven data quality. Such an account leaves out the structural features of strategic decisions themselves.

The risk originates in those features of strategic decisions, not in any general property of the technology. In cutting-edge research on firms, there has been a focus on capability – studies have revealed that large language models can create and evaluate

© The Author(s) 2026
B. P. Siba et al. (eds.), *Proceedings of the 2026 5th International Conference on Public Service, Economic Management and Sustainable Development (PESD 2026)*, Advances in Economics, Business and Management Research 411,
https://doi.org/10.2991/978-94-6239-825-2_27

strategies in a way that rivals the capabilities of entrepreneurs and investors, improving the speed, quality, and scale of strategic analysis [1]. The risk perspective has less to say. Existing research remains fragmented, and little attention has been given to one-off decisions that cannot be measured and are irrevocable.

This particular framework is not designed for such a situation. The COSO framework for enterprise risk management is a model that links risk management with strategy and performance through five components and twenty principles, [2] and companies operate this framework alongside the traditional approach of assurance responsibility division. Both share the same assumption applicable to most organizational decisions: there are grounds upon which the decision can be made, the problem has been framed before assessing risks against it, there is someone to assign responsibility, and risk comes from within the organization itself. When embedding a model in a strategic decision, all four assumptions are disturbed, and the degree of disturbance depends on the depth of the embedding process.

Two questions arise. What are the risks of integration, and how does layering affect them? How does layering change when AI shifts from providing information to making decisions? The conceptual analysis developed here answers with a three-tier hierarchy consistent with the responsibilities that businesses already have, along with an implication that arises out of this alignment; the tier that carries unlimited risk exposure is the one without any assurance lines.

2 Literature Review

2.1 AI in Strategic Decision-Making

AI does not come into a strategic decision in a unified form; rather, it comes into a decision at a certain depth, and that is what determines what may fail. The difference between automation, which involves a machine taking over an operation, and augmentation, which entails humans working with machines, supplies the starting vocabulary. A paradoxical interpretation of the difference suggests that they are inseparable from each other in the context of management, since they remain interdependent across time and space and emphasis on either sets off reinforcing cycles [3]. The implication runs counter to any such neat schema, including the current one. If automation and augmentation are intertwined, then how far a particular model is involved in a given decision does not depend on something that a company sets and sticks to; rather, it changes based on practice, load, and recommendation strength.

That instability is the reason to name positions at all. Informing is where AI presents input for the person to process; recommending is where AI ranks choices or gives recommendations; deciding is where AI makes an output which is directly executed. The three mark stations along a path rather than states through which a company travels. They are important to us because of the different control of the final decision that exists at each point. The position of a firm at any one time may be uncertain, but its direction is not.

The three positions are therefore analytical anchors on a continuum of AI decision authority rather than an exhaustive classification of organizational deployment patterns.

A single strategic process may move repeatedly between them: managers may revise assumptions and request new outputs, override a recommendation, or escalate an automated output for committee approval. Iterative feedback describes repeated exchanges between human users and the system; human override describes the substantive authority to reject or reverse an output; exception escalation transfers low-confidence or high-impact cases to a higher level; and distributed decision-making allocates review across business, technical, risk, and governance actors. These patterns do not replace the three positions. They explain how human authority is exercised within each position and why nominally similar deployments may produce different exposure.

Past research tends to focus on the middle node, and so too does the analysis that follows in the text below. It turns out that reliance is malleable and not fixed because of an experiment conducted with about 1,500 people that showed that performance pay and tournament pay increase reliance compared to fixed pay [4].

2.2 Enterprise Risk Management and the Gap

Enterprise risk management provides the framework, and its adaptation for AI is currently underway. The COSO principles have been adapted to AI projects, with recommendations that thresholds in performance and risk metrics should result in off-cycle analysis of models, whose results should be communicated to senior executives [5]. Public standards have evolved similarly through a voluntary, domain-neutral model that classifies risk management tasks as governance, mapping, measurement, and management across the system lifecycle [6]. Responsibility is conventionally divided across three lines: business ownership, risk and compliance oversight, and internal audit.

There are four premises that make architecture work, and AI on top disturbs each in a different place. The premise for the basis of decision-making is that it should be explainable; an unexplainable model cuts the connection between the recommendation made and its review. Risk is then assessed against objectives already set, which presumes the framing is settled before assessment begins. This assumption does not hold when the device used in defining the problem can optimize only within the boundaries of its own frame. Attribution requires the presence of an author; however, the decision trace combining the input of humans and machines does not have one. And exposure is considered an internal phenomenon, an assumption which breaks down as soon as capabilities come from the external suppliers. The stations mentioned above represent a distinct dimension, cutting through the layers described below.

The fourth disruption has finally drawn regulatory attention. Out of 184 market participants surveyed, nearly half have adopted AI, while the report that accompanied the survey identifies concentration, outsourcing, and reliance on third parties as key sources of risks [7]. Guidance and funding continue to be fine-tuned for practical and generalized uses, where the results can be repeated, measured against actual results, and fixed for the next round. Strategy-making provides none of these, and the result that comes after that point is undefined.

3 A Three-Layer Risk Framework

The disturbances mentioned above indicate that there is a stratified structure instead of the linear one where risks are concerned. Model opacity and problem definition lock-in affect the production of the decision, whereas attribution uncertainty and external dependency affect accountability. Two of the emerging layers fit into the assurance lines easily, but not the third one, and this misfit is telling (see Table 1).

Table 1. The three-layer risk structure and its assigned responsibilities.

Layer	Nature of risk	Where responsibility sits
Technical	Defects in the data and the model	Business ownership
Decisional	Judgement distorted in human-ma chine interaction	Risk and compliance over-sight
Institutional	Accountability, compliance, external de-pendency	Board and executive govern-ance

Accountability, risk tolerance, and supplier dependence ultimately lie with the governing body. The institutional layer is therefore the hardest to place and the most resistant to controls designed for the other two layers. Two quantities must be distinguished. Absolute exposure denotes the level of risk within a layer, whereas relative exposure denotes that layer's share of the total risk profile. As AI decision authority increases, absolute technical exposure may rise before approaching the ceiling imposed by validation and process controls, even while its relative share declines because institutional exposure grows more rapidly. Decisional exposure is expected to be greatest at recommendation, where human judgment remains formally present but is vulnerable to anchoring. Institutional exposure is expected to increase throughout because AI outputs become more directly connected to organizational action.

Proposition 1a. As AI decision authority increases, absolute technical exposure increases and then approaches a control-dependent ceiling, although its relative share of total exposure may decline.

Proposition 1b. Decisional exposure is expected to reach its highest level when AI recommends a course of action while humans retain formal decision authority.

Proposition 1c. Institutional exposure is expected to increase as AI outputs become more directly connected to organizational action.

Across this progression, three established enterprise risk management tools lose traction: control testing requires a verifiable outcome, ex post audit requires an articulable rationale, and exposure allocation requires an identifiable owner. Problem framing, the fourth disturbance, was not a primary object of these controls (see Fig. 1).

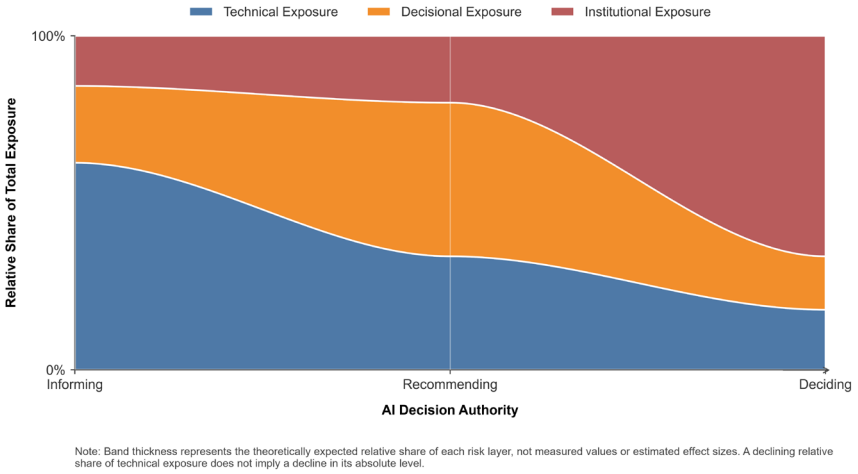


Fig. 1. Conceptual composition of AI risk exposure.

The proposed patterns receive different levels of support from prior evidence. Experimental research involving approximately 1,500 participants shows that incentives and framing affect reliance on algorithmic advice [4], supporting the mechanism behind heightened decisional exposure at recommendation. Research on six manufacturing firms documents gradual AI involvement in international strategic decision-making and continuing human caution , consistent with treating informing, recommending, and deciding as positions rather than fixed organizational states. The IOSCO survey, which received 184 responses, identifies concentration, outsourcing, and third-party reliance as material AI risks [7], supporting the institutional mechanisms. These studies support components of the framework; because none directly compares all three layers across all three positions, they should not be interpreted as conclusive validation of the complete trajectories.

A firm managing AI risk chiefly through model validation has fortified the layer whose share is falling, while the exposure it has not equipped for accumulates above.

4 Layered Analysis

4.1 Technical Layer

Strategic decision-making occurs very infrequently and is almost never repeated. It is these traits that render most of what model risk management ordinarily does ineffective.

Lack of training examples and absence of similar precedents make the model predict far out-of-distribution compared to what it was trained on. The harder problem is the missing target: without a realized outcome to score against, backtesting and cross-validation have nothing to test.

Such a testing methodology fails to satisfy these two conditions. It is periodic and takes place against a static specification, while also assuming that the correctness of its

output can be proved. The only thing left is indirect. Data lineage tracking ensures accountability for the input data even if the output data are not. Given the lack of outcome metrics within any timeframe useful for monitoring, process metrics have to be used.

Involvement controls the amount that gets through to the decision process. The indirect proxies outlined above will continue to work all along, providing an upper bound up to which the exposure can rise and then level off.

4.2 Decisional Layer

The condition of ambiguity is fundamental here. Strategic problems come with no preexisting definition, their framing is ambiguous, and figuring out what to inquire about can be the real question. Executive attention is finite and rationed.

The governance tool that the company turns to is human review. It succeeds in situations where the person conducting the review possesses the capacity to assess the recommendation, the motivation to overturn the recommendation, and the time to do so. This means that at least one condition for proper review is built into the firm's remuneration scheme. Remove even one of the three, and review degrades into formal compliance while still conferring the appearance of oversight, consuming assurance capacity that a working control would have required.

Even when complete, review does not reach the exposure that matters most here. Framing lock-in is immune to review. What it cannot do is ask whether those criteria capture what matters, or whether the decision-making question has itself been framed correctly. If the strategic problem at hand involves answering this question, then delegating the answer while leaving the framing unexamined narrows the decision space before deliberation begins. The narrowing is difficult to detect precisely because what arrives is fluent, internally consistent, and complete. A reviewer checking the recommendation against the criteria will find nothing wrong.

Other distortions run parallel to this one. Automation bias is the tendency to be excessively confident in automated results despite the signals from others which indicate otherwise. False precision does comparable damage, since a ranked list with numerical scores displaces the qualitative judgement that ambiguous problems require. There are two solutions here which close the gap: recording at which points the human decision went against the recommendation made and the rationale for doing so, and adversarial framing tests that oblige the model to produce alternative definitions of the problem.

The exposure here reaches its peak at recommendation and diminishes from there on out, since a decision executed without substantive human revision leaves no judgement left to distort. The human at the middle station remains nominally responsible while being substantively anchored, and this analysis takes that position as its focus because it is where review is formally present and least effective.

4.3 Institutional Layer

Irreversibility defines this layer. A strategic decision sets direction, and the organization cannot run the counterfactual to learn what the alternative would have produced.

The minute the outcome from the algorithmic process finds its way into the decision process, the chain of accountability is broken in such a manner that no subsequent examination can put things right. The obligation to comply becomes meaningless at the very instant. The modern governance regime requires companies to classify systems based on risks, maintain documentation throughout the life cycle, and assess systems prior to deployment [7]; each requirement is linked to an owner who is identified, and where ownership is indeterminate, the requirement remains formally in force while resisting operationalization.

The second source of vulnerability is supplier concentration. The capabilities, pricing, and continued availability of the base products are controlled by only a few outside suppliers [8], meaning that the firm's strategic analytics capacity is based on business decisions over which it has no control. Neither of these exposures is the type of risk the assurance lines were set up to capture. The lines own, control, and assure activity that the firm does. An indeterminate decision record is an internal artifact that has ceased to support any attribution, while the structure of an external market is not the firm’s activity.

Only structural remedies will do. The appointment of an ownership governance committee within the board will ensure that the layer can be a clear owner of the exposure to the model. The risk appetite should be viewed as non-delegable. It is the role of the governing body under current governance practices, not the instrument used.

Think about capital allocation that has been reviewed two years later and where returns were less than expected. In such a case, there are records showing that there was a ranking from the model, revised shortlist, and approval. Establishing which of the three was at fault is not a matter of investigative diligence; the record no longer contains the distinction.

Institutional exposure increases across the full continuum, and existing assurance-line controls do not impose an effective ceiling on it. The required limitation must therefore be established through governance arrangements. Enhanced model validation strengthens the technical layer but cannot resolve responsibility that remains indeterminate.

4.4 Foundation Models and Generative AI

Foundation models intensify rather than replace the three layers. At the technical layer, confabulation, prompt sensitivity, non-deterministic generation, uncertain source provenance, and rapid model-version changes make a fluent output difficult to treat as a stable decision input [8,9]. The magnitude of prompt sensitivity varies across models, tasks, and evaluation procedures; it is therefore a context-dependent exposure rather than a universal constant. Technical controls should include source verification, prompt and output logging for material decisions, model-version records, and re-evaluation after material updates.

At the decisional layer, generative systems can participate not only in ranking alternatives but also in defining the question, generating the alternatives, and presenting a persuasive rationale. Fluency may conceal unsupported claims, while a generated explanation is not necessarily a verifiable account of the model’s internal reasoning.

Multi-turn interaction can also shift a system between informing and recommending without a formal change in its deployment label. For irreversible decisions, generated claims and citations therefore require independent checking, and reviewers must retain the time, competence, and authority to challenge both the recommendation and its framing [10].

At the institutional layer, reliance on externally supplied foundation models deepens supplier concentration and divides responsibility among the model provider, deploying organization, prompt designer, reviewer, and final decision-maker. Model updates may alter behavior without a corresponding change in internal governance. Firms should therefore establish accountable ownership, model-change controls, confidentiality rules, a supplier exit plan, and a presumption against direct execution of irreversible strategic decisions by a general-purpose generative model [11], which is summarized in Table 2.

Table 2. Summary of the three-layer risk framework.

Dimension	Technical	Decisional	Institutional
Strategic-decision feature	Low frequency; no ground truth; delayed feedback	Contested framing; limited executive attention	Irreversibility; long-term organizational consequences
Primary exposure	Out-of-distribution inference; validation failure; undetected drift	Framing lock-in; over-reliance; false precision	Accountability failure; supplier concentration
ERM limitation	Control testing requires a verifiable outcome	Ex post audit requires an articulable rationale	Existing assurance lines cannot fully address distributed responsibility and external dependency
Governance response	Data lineage; process indicators; independent validation	Meaningful human review; divergence logging; alternative framing tests	Board-level ownership; non-delegable risk appetite; supplier oversight
Expected trajectory	Absolute exposure rises and then approaches a control-dependent ceiling; relative share may decline	Peaks at recommending	Rises with AI decision authority
Supporting evidence	Six-SME study and NIST guidance	Algorithmic-advice experiment	IOSCO survey of 184 respondents

The cited evidence supports the underlying mechanisms but does not directly validate the complete exposure trajectories.

5 Implications and Conclusion

5.1 Managerial Implications

Reversibility, and not weight, is what should define the extent of application for AI technology. The former is about the risk of something going wrong that can be addressed within the time frame of corporate planning while the latter is about how much

there is to lose. A reversible decision is one that allows a model to be sometimes inaccurate. An irreversible one is not.

Decisions with low reversibility should ordinarily remain closer to informing and should not move toward deciding merely because a model has performed well on comparable tasks. Greater authority may be defensible only where the firm can demonstrate meaningful human override, traceability, independent challenge, and board-level accountability. Comparability is precisely the assumption that the technical layer cannot guarantee: performance on decisions taken monthly establishes little about a decision taken once.

When there is an increase in involvement, the degree of accountability needs to be increased accordingly. As per the analysis above, exposure without ceiling is at the institutional level; when a firm increases its level of involvement without increasing ownership at the institutional level, it is widening the very gap that the framework was designed to highlight.

Outcome metrics are simply not available within any time frame that could be useful for making strategic decisions; therefore, surveillance must rely on process metrics, and should be embedded in the existing risk-reporting process cycle of an organization rather than conducted in a separate technical activity. A model review prompted by unfavorable outcomes may never materialize at all.

5.2 Conclusion, Limitations, and Future Research

AI integration creates technical, decisional, and institutional exposure whose importance changes with decision authority. The available experimental, organizational, and regulatory evidence is broadly consistent with the mechanisms proposed here, although it does not conclusively validate the complete trajectories. Technical exposure can be constrained but not eliminated by model and process controls; decisional exposure is especially salient when human review is formally retained but substantively weakened; and institutional exposure requires explicit governance ownership as AI outputs become more directly connected to action. Enterprise risk management therefore does not require an entirely separate architecture, but it does require responsibility, reversibility, and meaningful human authority to be recalibrated as AI involvement deepens.

Several caveats apply. The framework remains conceptual, and the evidence summarized in this article supports individual mechanisms rather than directly estimating the three complete trajectories. The three layers are analytically separable but interact in practice; for example, technical opacity can produce institutional accountability failures. Informing, recommending, and deciding are anchors on a continuum of decision authority, not an exhaustive taxonomy of iterative, override-based, escalated, or distributed human–AI configurations. The organizational illustrations rely on previously published studies and public regulatory evidence and therefore demonstrate applicability rather than causal effectiveness. Foundation-model capabilities and supplier practices also change rapidly, so controls must be revisited after material system updates. Direct comparative research across the three positions remains an important next step.

There is one challenge that might remain unanswered. The calibration between involvement and accountability requires that the company is able to know where along the spectrum the decision falls. Whether this can be done beforehand or whether the decision will be reviewed afterward to see its place along the spectrum, is a challenge this model poses but does not answer.

References

1. Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. Strategy Science, 9(4), 322–345. https://doi.org/10.1287/stsc.2024.0190.
2. Committee of Sponsoring Organizations of the Treadway Commission. (2017). Enterprise risk management: Integrating with strategy and performance. COSO.
3. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review, 46(1), 192–210. https://doi.org/10.5465/amr.2018.0072.
4. Greiner, B., Grünwald, P., Lindner, T., Lintner, G., & Wiernsperger, M. (2026). Incentives, framing, and reliance on algorithmic advice: An experimental study. Management Science, 72(1), 302–322. https://doi.org/10.1287/mnsc.2022.02777.
5. Calagna, K., Cassidy, B., & Park, A. (2021). Enterprise risk management: Realize the full potential of artificial intelligence: Applying the COSO framework and principles to help implement and scale artificial intelligence. Committee of Sponsoring Organizations of the Treadway Commission.
6. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1.
7. International Organization of Securities Commissions. (2025). Artificial intelligence in capital markets: Use cases, risks, and challenges (Consultation Report CR/01/2025). IOSCO. https://www.iosco.org/library/pubdocs/pdf/IOSCOPD788.pdf.
8. Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). *National Institute of Standards and Technology*. https://doi.org/10.6028/NIST.AI.600-1.
9. Zhuo, J., Zhang, S., Fang, X., Duan, H., Lin, D., & Chen, K. (2024). ProSA: Assessing and understanding the prompt sensitivity of LLMs. Findings of the Association for Computational Linguistics: EMNLP 2024, 1950–1976. https://doi.org/10.18653/v1/2024.findings-emnlp.108.
10. Cheng, J., Su, T., Yuan, J., He, G., Liu, J., Tao, X., Xie, J., & Li, H. (2025). Chain-of-thought prompting obscures hallucination cues in large language models: An empirical evaluation. Findings of the Association for Computational Linguistics: EMNLP 2025, 1272–1305. https://doi.org/10.18653/v1/2025.findings-emnlp.67.
11. Organisation for Economic Co-operation and Development. (2023). Generative artificial intelligence in finance. *OECD Publishing*. https://doi.org/10.1787/ac7149cc-en.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

