

Article

Not peer-reviewed version

An Agent Based Framework for Assessing Risk Actionability in SEC Filings

[Weishen Chu](#) *

Posted Date: 29 December 2025

doi: 10.20944/preprints202512.2509.v1

Keywords: interpretable AI; rule-based agents; text-based risk assessment; human-in-the-loop systems; decision-centric modeling



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

An Agent Based Framework for Assessing Risk Actionability in SEC Filings

Weishen Chu

Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, USA;
weishenchu2014@u.northwestern.edu

Abstract

Risk factor disclosures in corporate filings are widely analyzed for their informational content, yet their usefulness for operational governance remains underexplored. This paper introduces risk actionability as a distinct concept from risk presence and proposes an interpretable, agent-based pipeline for assessing whether disclosed risks support intervention-oriented decisions. Operating at the paragraph level on SEC Form 10-K and 10-Q risk factors, the agent assigns risk attributes and maps disclosures to pass, review, or escalate states. We introduce the Intervention Load Index and Actionability Gap to quantify human oversight demand implied by disclosures. Empirical results show that disclosure severity does not reliably predict intervention needs, and that conservative disclosure styles can suppress actionable signals. A failure case highlights limitations arising from disclosure sparsity.

Keywords: interpretable AI; rule-based agents; text-based risk assessment; human-in-the-loop systems; decision-centric modeling

1. Introduction

Risk factor disclosures in U.S. public company filings are a central channel through which firms communicate uncertainties to regulators, auditors, and investors. In Form 10-K, issuers are required to disclose material risk factors under Item 1A, while Form 10-Q focuses on material changes relative to the most recent annual disclosure, making these sections a standardized and recurring corpus of narrative risk statements across firms and time [1–3]. Prior research has shown that narrative text in 10-K filings contains economically meaningful signals and can be systematically analyzed at scale, motivating a growing literature on computational analysis of financial disclosures [4]. Studies of mandated risk factor disclosures further suggest that such statements can be firm-specific and informative rather than purely boilerplate [5].

Despite this progress, most existing work evaluates risk disclosures in terms of information content, market reactions, or predictive power, rather than their usefulness for operational governance. Textual risk measures have been linked to outcomes such as volatility, returns, and litigation risk [6,7], yet these analyses largely treat risk disclosure as an input to statistical inference rather than as a trigger for concrete oversight actions. In regulatory and auditing contexts, however, the practical question is not only whether risks are disclosed, but whether the disclosed statements meaningfully support intervention decisions such as escalation, targeted review, or dismissal. As a result, extensive or cautious disclosures may appear informative while still failing to produce actionable signals for governance.

In this paper, we introduce *risk actionability* as a distinct concept from risk presence and study whether disclosed risks support intervention-oriented decision making. We propose a transparent agent-based pipeline that processes SEC 10-K and 10-Q risk factor texts at the paragraph level, assigns risk attributes including severity and type, and maps each paragraph to one of three decision states: PASS, REVIEW, or ESCALATE. The agent logic is intentionally simple and rule-based, allowing decision outcomes to be directly audited and interpreted. Building on agent outputs, we introduce two

metrics—Intervention Load and Actionability Gap—to quantify how disclosure structure translates into governance-relevant actions rather than abstract risk indicators.

Using risk factor disclosures from major technology firms, we demonstrate that disclosure intensity and actionability can diverge substantially. Firms with more extensive or cautious disclosures do not necessarily exhibit higher intervention demand, and conservative disclosure styles may suppress agent-driven differentiation across risk statements. A failure case study on NVIDIA further illustrates a disclosure-limited regime in which risk statements are highly abstract and forward-looking, resulting in sparse actionable signals despite the presence of material uncertainties. Together, these findings highlight that actionability is jointly shaped by disclosure practices and decision logic, and cannot be inferred from risk volume or severity alone.

2. Related Work

2.1. Textual Analysis of Financial Disclosures

A large body of prior work studies the informational content of financial disclosures using textual analysis. Early research demonstrates that narrative sections of corporate filings, particularly Form 10-K, contain economically meaningful signals beyond traditional numerical indicators [4]. Subsequent studies develop domain-specific dictionaries and statistical models to extract measures of risk sentiment, uncertainty, and tone, and link them to outcomes such as stock returns, volatility, and firm performance [8]. More recent work applies machine learning and natural language processing techniques to predict firm-level risk and market reactions directly from disclosure text [6]. While these approaches establish the value of textual risk analysis, they primarily treat disclosure text as input for prediction or inference rather than as a basis for operational decision making.

2.2. Risk Factor Disclosures and Regulatory Context

Another line of research focuses specifically on mandated risk factor disclosures and their regulatory implications. Studies show that risk factor statements under Item 1A of Form 10-K are informative, firm-specific, and associated with litigation risk and information asymmetry [5]. Related work examines how firms update risk disclosures over time and how such changes reflect evolving risk exposures [9]. These studies emphasize compliance and informativeness, but generally evaluate disclosures in terms of completeness, novelty, or market impact. Less attention has been paid to whether the structure and phrasing of risk disclosures support downstream governance actions such as escalation or targeted review by regulators and auditors.

2.3. Agents and Decision Support in Financial Text Analysis

Recent research explores automated decision support systems and agents for financial text analysis, including applications in auditing, compliance, and regulatory technology. Prior work highlights the importance of transparency and interpretability in such systems, particularly in high-stakes financial settings [10]. Rule-based and weakly supervised approaches are often favored in these contexts due to their auditability and alignment with institutional processes [11]. However, existing systems typically focus on classification or anomaly detection tasks rather than on explicitly modeling intervention-oriented decision states. In contrast to predictive or black-box approaches, this paper examines how simple, interpretable agent logic interacts with disclosure practices to shape the actionability of risk statements.

3. Methodology: Agent-Based Risk Assessment Pipeline

This section presents an agent-based pipeline designed to assess the actionability of corporate risk disclosures at the paragraph level. The pipeline transforms unstructured textual disclosures into structured decision signals that indicate whether human intervention is required. An overview of the proposed pipeline is shown in Figure 1.

Figure 1. Agent-Based Pipeline for Risk Actionability Analysis

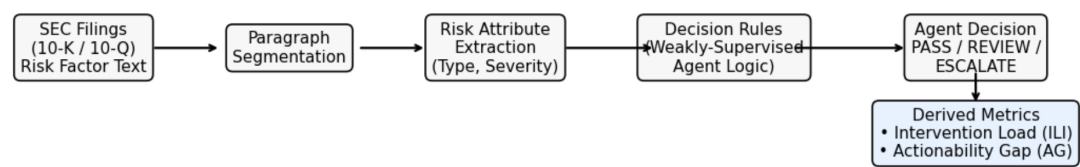


Figure 1. Agent-based pipeline for risk disclosure analysis. The pipeline segments SEC risk disclosures into paragraph-level units, assigns severity and risk type attributes, maps each paragraph to an intervention decision (PASS, REVIEW, ESCALATE), and aggregates decisions into firm-level actionability metrics.

3.1. Pipeline Overview

The pipeline consists of four sequential stages: paragraph segmentation, risk attribute annotation, agent decision-making, and document-level aggregation. Each stage is modular and interpretable, enabling transparent analysis of how raw disclosures are transformed into intervention signals.

Given a corporate disclosure document (e.g., an annual risk factors section), the text is first segmented into semantically coherent paragraphs. Each paragraph is then independently processed by the agent to avoid conflating heterogeneous risk statements. The agent assigns structured risk attributes to each paragraph and produces a discrete decision indicating whether human oversight is required.

3.2. Paragraph-Level Processing

Paragraphs are chosen as the unit of analysis because they typically correspond to a single risk theme or conditional statement in regulatory disclosures. Compared to sentence-level processing, paragraph-level analysis preserves contextual coherence, while avoiding the ambiguity and dilution that arise in document-level classification.

Each paragraph is treated as an independent input instance, allowing the agent to generate fine-grained intervention signals that can later be aggregated across the document.

3.3. Risk Attribute Annotation

For each paragraph, the agent assigns two categorical attributes:

- **Risk Severity**, representing the potential impact of the disclosed risk. Severity is discretized into three ordered levels: *Low*, *Medium*, and *High*.
- **Risk Type**, representing the thematic category of the risk. Examples include regulatory and legal risk, operational risk, technological risk, financial risk, and human capital risk.

These attributes are derived from the content of the paragraph and are intended to capture how serious the risk is and what domain it pertains to, without relying on quantitative financial forecasts.

3.4. Agent Decision Logic

Based on the annotated risk severity and risk type, the agent produces a discrete intervention decision with three possible outcomes:

- **ESCALATE**: The paragraph indicates a risk that requires mandatory human intervention.
- **REVIEW**: The paragraph warrants secondary inspection or sampling-based review.
- **PASS**: The paragraph does not require immediate human attention.

The decision logic is rule-based and deterministic, designed to be transparent and auditable. High-severity risks are always escalated, while medium-severity risks may trigger review depending on their thematic category. Low-severity risks are typically passed without intervention. This design prioritizes interpretability and governance alignment over predictive optimization.

3.5. Output Representation and Aggregation

For each document, the agent produces a set of paragraph-level annotations and decisions. These outputs can be aggregated to derive document-level statistics, such as severity distributions, decision mixes, and intervention rates. Importantly, aggregation is performed only after paragraph-level decisions are finalized, ensuring that individual risk signals remain traceable.

The resulting structured outputs enable systematic comparison across documents and organizations, and form the basis for the empirical analysis presented in subsequent sections.

4. Results

This section presents empirical results obtained by applying the proposed agent-based pipeline to corporate risk disclosures. We evaluate the distribution of annotated risk severity, agent decision outcomes, and derived intervention signals across firms. All experiments are conducted at the paragraph level using risk factor disclosures from SEC filings, consistent with prior work on regulatory text analysis [? ?].

4.1. Severity and Risk Structure Analysis

We first examine the distribution of annotated risk severity levels for Apple Inc. (AAPL) and Micron Technology (MU). Figure 2 compares the proportion of paragraphs classified as Low, Medium, and High severity for each firm.

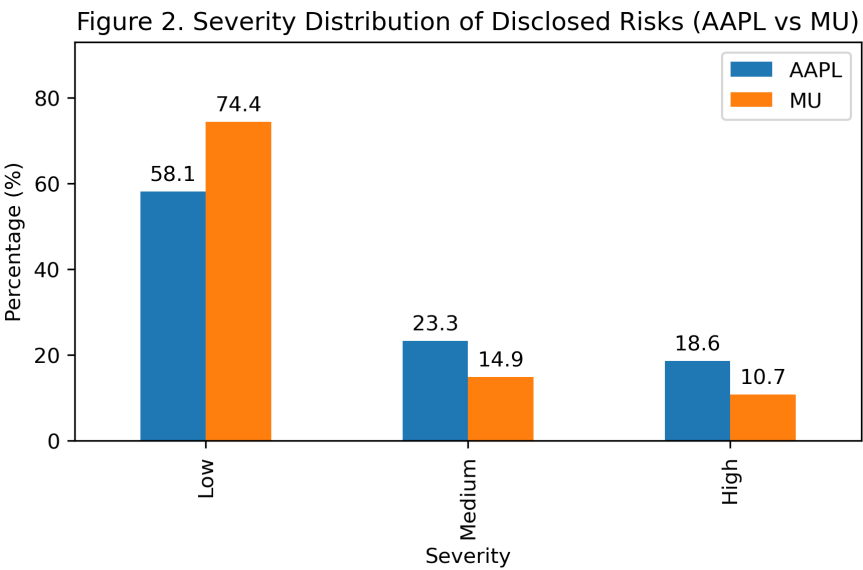


Figure 2. Severity distribution of disclosed risks for AAPL and MU.

As shown in Figure 2, both firms predominantly disclose risks categorized as Low severity. However, AAPL exhibits a noticeably higher share of Medium and High severity statements compared to MU. Specifically, High severity disclosures account for a larger fraction of AAPL’s risk paragraphs, suggesting a more granular articulation of potential adverse outcomes. In contrast, MU’s disclosures are more heavily concentrated in the Low severity category, reflecting a more conservative or generalized disclosure style.

These differences highlight that severity distributions alone do not directly translate to intervention demand, motivating further analysis of agent decision outcomes.

To analyze how risk categories interact with severity, Figures 3 present heatmaps of risk type versus severity for AAPL and MU, respectively.

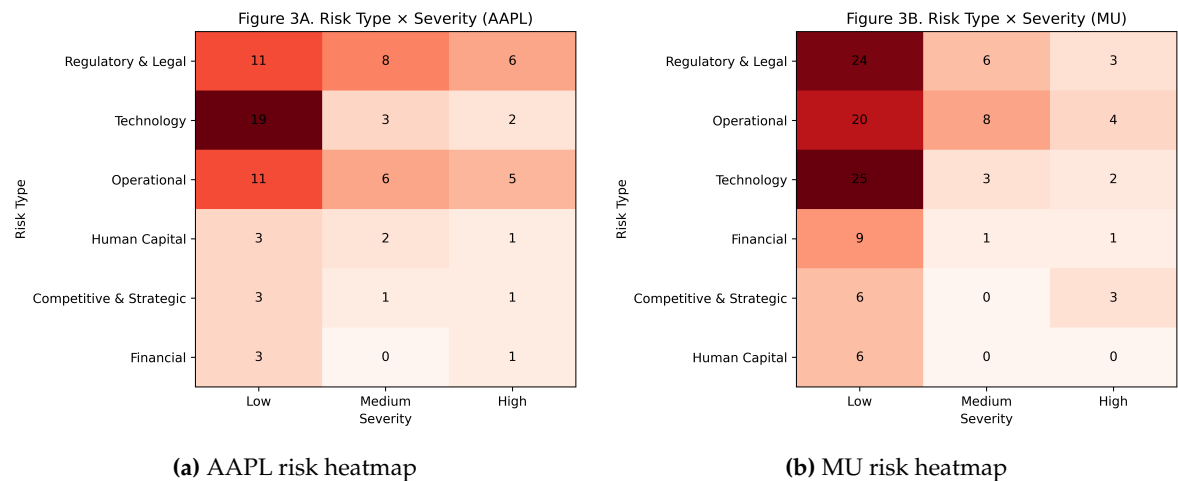


Figure 3. Risk type by severity heatmaps for AAPL and MU.

For both firms, technology and regulatory risks dominate the disclosure volume. AAPL demonstrates a broader dispersion across severity levels within these categories, including non-trivial counts of Medium and High severity statements. MU, by contrast, concentrates most risk types at Low severity, with relatively sparse High severity disclosures across categories.

These patterns suggest that AAPL provides more differentiated severity signaling across risk types, while MU’s disclosures emphasize breadth over depth.

4.2. Agent Decision Outcomes and Intervention Patterns

Figure 4 summarizes the agent’s paragraph-level decisions for both firms. Decisions are categorized as PASS (no human intervention), REVIEW (selective human verification), or ESCALATE (mandatory human intervention).

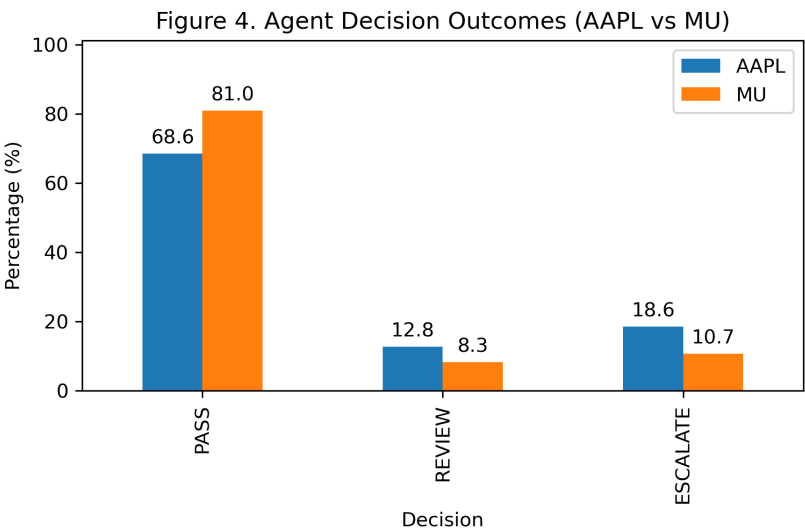


Figure 4. Agent decision outcomes for AAPL and MU.

Despite AAPL disclosing a higher proportion of Medium and High severity risks, the agent escalates a larger share of AAPL paragraphs compared to MU. MU exhibits a higher PASS rate, with fewer paragraphs flagged for REVIEW or ESCALATE.

This outcome indicates that higher disclosure severity does not mechanically imply higher intervention demand. Instead, the interaction between severity, risk type, and agent decision logic shapes the resulting intervention signals.

To further illustrate intervention patterns, Figure 5 presents the stacked decision mix for each firm. AAPL exhibits a higher proportion of REVIEW and ESCALATE decisions relative to its Low severity

disclosure share. This suggests that many paragraphs classified as Low severity still warrant human attention due to contextual or categorical considerations. MU, by contrast, shows closer alignment between Low severity disclosures and PASS decisions.

Figure 5. Decision Mix (PASS / REVIEW / ESCALATE) by Company

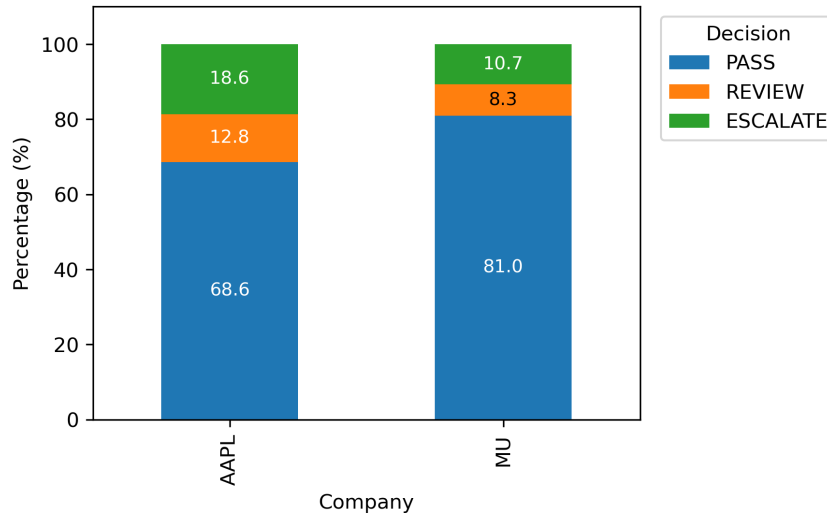


Figure 5. Decision mix (PASS / REVIEW / ESCALATE) by company.

These results reveal a systematic divergence between disclosure severity and intervention demand, underscoring the importance of modeling actionability explicitly rather than inferring it from severity alone.

4.3. Intervention Load and Actionability Gap

To quantify the operational burden implied by agent decisions, we define the *Intervention Load Index* (ILI) as the fraction of paragraphs requiring human attention:

$$ILI = \frac{N_{REVIEW} + N_{ESCALATE}}{N}. \quad (1)$$

Because ESCALATE typically incurs higher human cost than REVIEW, we also report a severity-weighted variant:

$$ILI_{sev} = \frac{N_{ESCALATE} + 0.5 \cdot N_{REVIEW}}{N}. \quad (2)$$

To characterize the mismatch between disclosure severity and intervention outcomes, we further define the *Actionability Gap* (AG) as:

$$AG = \%PASS - \%Low. \quad (3)$$

A larger AG indicates that a substantial fraction of Medium- or High-severity disclosures are still auto-passed by the agent, whereas an AG close to zero suggests closer alignment between Low severity labeling and automated decisions.

Table 1 reveals that AAPL imposes a substantially higher intervention burden than MU under the same transparent decision logic, with $ILI_{sev} = 0.25$ for AAPL compared to 0.149 for MU. Notably, AAPL also exhibits a larger Actionability Gap (AG), indicating that a non-trivial portion of Medium- and High-severity paragraphs are still auto-passed by the agent. This finding reinforces that disclosure severity alone is insufficient to infer operational intervention demand, and that actionability emerges from the interaction between disclosure style, risk categorization, and decision policy.

Table 1. Intervention load and actionability gap across firms.

Company	N	Low (%)	PASS (%)	ILI	ILI _{sev}	AG
AAPL	86	58.1	68.6	0.314	0.250	10.5
MU	121	74.4	81.0	0.190	0.149	6.6
NVDA	100	98.0	98.0	0.020	0.010	0.0

4.4. Failure Case Analysis: NVDA

Finally, we analyze NVIDIA (NVDA) as a failure case. As shown in Figure 6, NVDA’s disclosures exhibit extremely low variance in severity and decision outcomes, with the vast majority of paragraphs classified as Low severity and auto-passed by the agent.

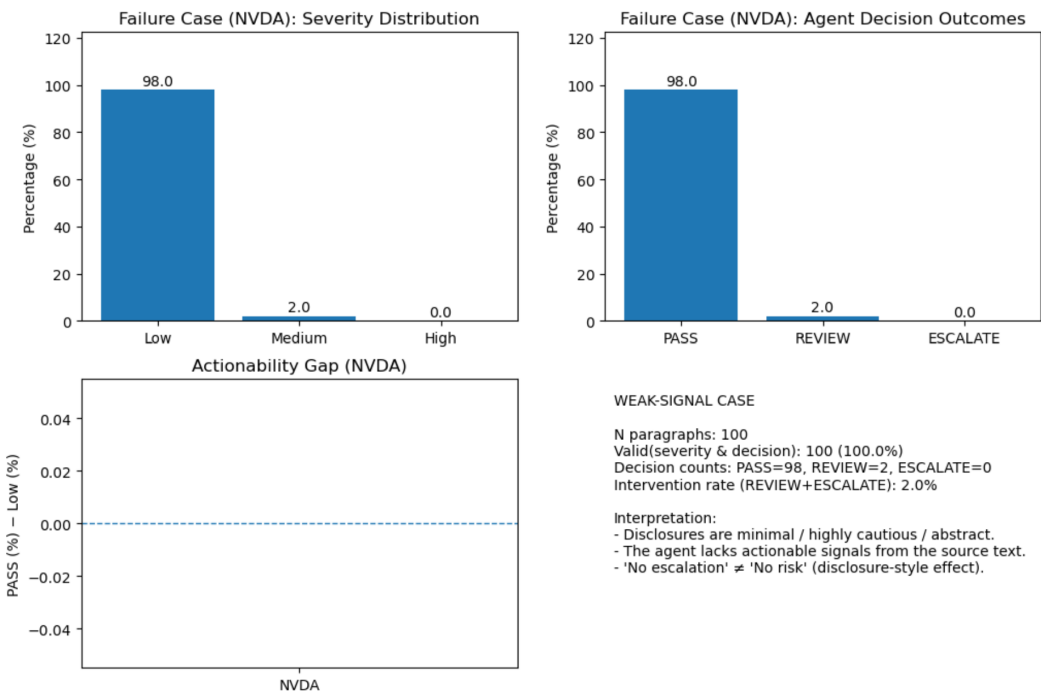


Figure 6. Failure case: NVDA exhibits disclosure-style sparsity that limits agent actionability.

This pattern reflects disclosure-style sparsity rather than the absence of operational or strategic risk. NVDA’s risk statements are highly abstract and cautious, limiting the availability of actionable signals at the paragraph level. Consequently, the agent produces minimal intervention recommendations.

This failure case highlights an important limitation of text-based actionability analysis: when disclosures intentionally minimize specificity, agent-based systems may underestimate intervention needs despite substantial underlying risk.

5. Conclusion

This paper proposes an interpretable, agent-based pipeline for assessing the actionability of corporate risk disclosures at the paragraph level. Rather than treating disclosure severity as a proxy for operational concern, the proposed framework explicitly models intervention-oriented decision states (PASS, REVIEW, ESCALATE) and aggregates them into document-level indicators of human oversight demand.

Through an empirical study of risk factor disclosures from multiple firms, we demonstrate that disclosure severity distributions alone do not reliably predict intervention requirements. In particular, Apple Inc. (AAPL) exhibits a higher intervention load than Micron Technology (MU) despite comparable dominance of low-severity disclosures. This divergence is captured quantitatively by the Intervention Load Index (ILI) and its severity-weighted variant, which reveal materially different

human review burdens under identical decision logic. The Actionability Gap (AG) further highlights systematic mismatches between low-severity labeling and automated pass decisions, underscoring that actionability is shaped by disclosure structure and contextual risk signals rather than severity annotations in isolation.

The failure case analysis of NVIDIA (NVDA) illustrates an important limitation of text-based actionability assessment. Highly abstract and cautious disclosure styles can suppress paragraph-level signals, resulting in minimal intervention recommendations even when substantive underlying risks are likely present. This finding emphasizes that low intervention load should not be conflated with low risk, and that disclosure practices themselves critically influence the effectiveness of automated oversight tools.

Overall, this work argues for a shift from severity-centric risk analysis toward intervention-aware evaluation of corporate disclosures. By providing transparent, auditable metrics such as ILI and AG, the proposed framework enables more faithful estimation of human oversight demand and offers a practical foundation for integrating agent-based systems into regulatory, auditing, and compliance workflows. Future work may extend this approach by incorporating temporal disclosure dynamics, cross-document consistency checks, and adaptive decision policies calibrated to institutional risk tolerance.

References

1. U.S. Securities and Exchange Commission. Form 10-K. <https://www.sec.gov/files/form10-k.pdf>. Accessed 2025-12-27.
2. U.S. Securities and Exchange Commission. Form 10-Q. <https://www.sec.gov/files/form10-q.pdf>. Accessed 2025-12-27.
3. Investor.gov. How to Read a 10-K and 10-Q. <https://www.investor.gov/introduction-investing/general-resources/news-alerts/alerts-bulletins/investor-bulletins/how-read>. Accessed 2025-12-27.
4. Loughran, T.; McDonald, B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance* **2011**, *66*, 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.
5. Campbell, J.L.; Chen, H.; Dhaliwal, D.S.; Lu, H.M.; Steele, L.B. The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings. *Review of Accounting Studies* **2014**, *19*, 396–455. <https://doi.org/10.1007/s11142-013-9258-3>.
6. Kogan, S.; Levin, D.; Routledge, B.R.; Sagi, J.S.; Smith, N.A. Predicting Risk from Financial Reports with Regression. *Proceedings of the Annual Meeting of the Association for Computational Linguistics* **2020**.
7. Hassan, T.A.; Hollander, S.; van Lent, L.; Tahoun, A. Firm-Level Political Risk: Measurement and Effects. *Quarterly Journal of Economics* **2019**, *134*, 2135–2202. <https://doi.org/10.1093/qje/qjz021>.
8. Loughran, T.; McDonald, B. Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research* **2016**, *54*, 1187–1230. <https://doi.org/10.1111/1475-679X.12123>.
9. Bao, Y.; Datta, S. Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures. *Management Science* **2014**, *60*, 1371–1391. <https://doi.org/10.1287/mnsc.2013.1800>.
10. Doshi-Velez, F.; Kim, B. Towards a Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608* **2017**.
11. Kraus, M.; Feuerriegel, S.; Oztekin, A. Deep Learning in Business Analytics and Operations Research: Models, Applications and Managerial Implications. *European Journal of Operational Research* **2020**, *281*, 628–641. <https://doi.org/10.1016/j.ejor.2019.09.018>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.