EIU Journal of Action Research

Intent-Aware AI for Early Risk Detection in Healthcare Payer–Provider Contracts: An FMEA-Based Governance Framework

Research-in-Progress Report

Prateek Gupta

Doctoral Candidate, Doctor of Business Administration (Generative AI)

Golden Gate University (GGU Worldwide)

prateekgupta1210@gmail.com

Submitted to: EIU Journal of Action Research

Abstract

This research-in-progress report presents a framework that combines large language models (LLMs), small language models (SLMs), Failure Mode and Effects Analysis (FMEA), and the ISO 31000 risk governance standard to detect and manage hidden risks in U.S. healthcare payer–provider contracts. Healthcare payer–provider contracts frequently contain ambiguous reimbursement language, prior authorization conditions, and appeal timelines that create financial and compliance exposure but remain difficult to detect before disputes arise. The framework identifies the intent behind ambiguous contract language—not just the clause classification. These intent interpretations are translated into structured risk signals, scored using FMEA's Risk Priority Number model, and mapped to ISO 31000 lifecycle stages to drive governance decisions. A proof-of-concept pilot on representative clause scenarios yields early evidence of feasibility. The framework aims to help contract lifecycle management (CLM) teams shift from reactive dispute resolution to proactive risk management. Specific calibration parameters will be reported in the complete dissertation. This report outlines the research problem, the approach, preliminary observations, and reflections on challenges encountered so far.

Keywords: *healthcare contracts, FMEA, intent-aware AI, ISO 31000, risk governance, contract lifecycle management, large language models*

1. Research Problem and Context

U.S. healthcare payer–provider contracts are complex legal instruments governing reimbursements, prior approval obligations, and denial resolution timelines. When these contracts function as intended, they are rarely scrutinized. When they fail, the consequences affect both financial performance and regulatory compliance. Contract mismanagement in U.S. healthcare contributes to over $157 billion in annual losses for providers (Black Book Market Research, 2019), and hospitals recover approximately 83 cents for every dollar spent under Medicare and Medicaid (American Hospital Association, 2023).

The financial risks embedded in healthcare contracts are not distributed evenly across clause types. Three categories account for a disproportionate share of disputes: fee schedule provisions, which create reimbursement gaps through vague adjustment language; prior authorization terms, which delay payments by leaving approval timelines undefined; and denial-and-appeal clauses, which impose narrow, ambiguously defined response windows that reduce recovery rates. In each case, the risk originates not from overt contractual error but from language that is technically compliant yet operationally exploitable.

Traditional Contract Lifecycle Management (CLM) systems support document digitalization and storage but provide limited capability for risk identification or governance. Current AI applications in contract management are primarily oriented toward clause extraction and keyword flagging rather than intent interpretation or structured risk quantification. No documented approach provides clause-level risk scoring with an auditable governance lifecycle. Industry evidence reinforces this gap: only 39% of organizations believe their contracts actively support business outcomes, 16% report that negotiations address the right risk issues, and 87% cite value leakage attributable to contract management weaknesses (World Commerce & Contracting [WorldCC], 2024a). This study proposes an AI-integrated framework that addresses these deficiencies through clause-level interpretation, FMEA-based risk scoring, ISO 31000 lifecycle alignment, and embedded Responsible AI safeguards.

2. Research Questions and Objectives

Four interconnected research questions structure the inquiry. RQ1 examines how FMEA and ISO 31000 can be adapted for clause-level risk modeling and governance in U.S. healthcare payer–provider contracts. RQ2 investigates how AI models can identify ambiguous clause-level language and generate FMEA-style risk scores. RQ3 considers how ISO 31000 lifecycle stages can be operationalized to create an auditable governance structure for AI-driven contract risk decisions. RQ4 addresses how Responsible AI principles—explainability, fairness, and privacy—can be embedded into the framework to ensure alignment with HIPAA and CMS requirements.

Together, these questions inform five research objectives. The first is to develop AI models capable of interpreting the intent behind ambiguous contract clauses. The second is to detect hidden financial and compliance risks at the clause level. The third is to apply FMEA-based risk scoring to produce quantified, comparable risk signals. The fourth is to align those signals with ISO 31000 treatment and monitoring stages within a structured governance lifecycle. The fifth is to embed Responsible AI safeguards calibrated to the healthcare regulatory environment, ensuring that framework outputs are explainable, fair, and privacy-preserving.

3. Background and Related Work

3.1 Healthcare Contract Risk Context

Contract risk varies across industries, but healthcare payer–provider agreements combine high financial exposure, structured legal language, and layered regulatory oversight. Unlike construction or software licensing contracts, healthcare agreements operate under frameworks including HIPAA, CMS reimbursement rules, and statutory billing requirements (Moon et al., 2022). Failures in these contracts create compounding financial and compliance consequences that extend across billing cycles.

3.2 From Clause Identification to Risk-Governed Automation

The literature on AI-based contract analysis can be understood through five interconnected themes that trace a progression from basic clause labeling to governance integration. CUAD (Hendrycks et al., 2021) provides 13,000 manually annotated clauses across 41 legal categories; LEDGAR (Tuggener et al., 2020) extends this to over 850,000 multi-labeled clauses from SEC filings. Legal-BERT (Chalkidis et al., 2020) and ALeaseBERT (Leivaditi et al., 2020) apply domain-specific transformer fine-tuning to produce category labels for each clause.

Document comprehension advances beyond categorization by modeling relationships between clauses. ContractNLI (Koreeda & Manning, 2021) enables entailment reasoning across clauses; ConReader (Xu et al., 2022) maps long-range semantic dependencies linking vague terms to their formal definitions within the agreement. PAKTON (Raptopoulos et al., 2025) extends this through a multi-agent design for cross-document policy retrieval. Purpose recognition further advances this progression: Bizzaro et al. (2024) extended CUAD to 33 clause types across 14 legal categories, identifying the legal purpose a clause serves rather than its structural type.

Risk quantification through textual analysis has been demonstrated in adjacent domains. Song et al. (2024a) derived ranked failure modes from hotel service complaints through LDA-based extraction and entropy weighting. Song and Zheng (2024) used BERT embeddings to automate FMEA Severity, Occurrence, and Detection scoring in auto parts processing. Sader et al. (2020) achieved F1 scores up to 0.89 predicting all three FMEA dimensions from engineering records. Collectively, these findings confirm that structured risk prioritization can be derived from unstructured textual data, yet the method has not been extended to legal contract language within a healthcare contracting context. Lifecycle integration similarly remains underdeveloped: ISO 31000 (International Organization for Standardization, 2018) defines four stages—Identify, Evaluate, Treat, and Monitor—yet reviewed AI contract studies address only Identify and partially Evaluate, leaving Treat and Monitor without structured AI support (Al-Mhdawi et al., 2023; Ittan, 2024; Wong et al., 2023).

3.3 Responsible AI in Legal Clause Automation

The legal AI literature has increasingly engaged with Responsible AI principles, though primarily through domain-specific implementations that address individual concerns in isolation. Explainability methods provide transparent reasoning for document review (Chhatwal et al., 2019; Mahoney et al., 2022). Fairness tools such as CLAUDETTE

(Lippi et al., 2019) and memory-augmented neural networks detect unfair clauses in consumer contracts (Ruggeri et al., 2021). Privacy-preserving approaches, including federated learning, protect sensitive legal data during analysis (Sonani et al., 2025). What the reviewed literature does not provide is an integrated RAI governance layer designed specifically for U.S. healthcare payer–provider contracting and calibrated to the combined requirements of HIPAA and CMS regulation—a gap the present framework directly addresses.

3.4 Comparative Analysis of Existing Approaches

Table 1 synthesizes five representative studies across four dimensions: contract scope, primary achievement, core limitation, and how this framework extends each boundary. The comparison illustrates where existing methods establish strong coverage and where they reach their boundaries relative to the present study's objectives.

Table 1

Comparative Analysis: Existing Approaches and Framework Extensions

Study	Scope	What It Achieves	Core Limitation	How This Framework Extends
Hendrycks et al. (2021) — CUAD	General commercial contracts	Expert-annotated clause identification across 41 legal categories; baseline for NLP contract research	Labels clause type only; does not assess operational risk, financial consequence, or governance action	Applies risk scoring to clause-level outputs, translating type labels into quantified risk signals with governance implications

Koreeda & Manning (2021) — ContractNLI	Non-disclosure agreements	Document-level entailment reasoning; evaluates logical consistency of contractual commitments	Assesses consistency; does not produce risk scores, financial exposure estimates, or governance outputs	Links intent interpretation to ISO 31000 Treat and Monitor stages, converting semantic findings into governance decisions
Song & Zheng (2024) — BERT-FMEA	Engineering and manufacturing records	Automates FMEA Severity, Occurrence, and Detection scoring from unstructured engineering text	Validated in manufacturing domain; not adapted for legal language or regulated contracting environments	Adapts the FMEA scoring mechanism to contractual ambiguity in U.S. healthcare payer–provider agreements under HIPAA and CMS constraints
Lippi et al. (2019) — CLAUDETTE	Consumer contracts (EU)	Detects unfair clauses using machine learning; fairness-focused classification for consumer protection	Fairness-only scope; does not integrate risk quantification, ISO governance, or healthcare regulatory requirements	Incorporates fairness detection alongside explainability and privacy within a unified RAI governance layer calibrated to CMS and HIPAA requirements

Ittan (2024) — AI Risk Scoring Engine	General contracts with AI-based risk focus	Develops an AI-driven contract risk scoring engine aligned with ISO 31000 Identify and Evaluate stages	Implements only Identify and Evaluate stages; does not operationalize Treat or Monitor; no healthcare-specific calibration or Responsible AI layer	Extends Ittan's scoring to the full ISO 31000 lifecycle (Treat, Monitor) with healthcare-specific calibration (HIPAA, CMS) and embedded Responsible AI safeguards

Collectively, the studies in Table 1 reveal four persistent gaps. First, clause-classification and intent-modeling research identifies what a clause says but does not score its operational or financial risk. Second, FMEA-based scoring has not been adapted for unstructured legal language in healthcare contracting. Third, no system implements the full ISO 31000 lifecycle—the most advanced AI approach (Ittan, 2024) covers only Identify and Evaluate. Fourth, Responsible AI safeguards remain siloed rather than embedded as cross-cutting constraints aligned with HIPAA and CMS requirements. These gaps define the contribution boundary addressed in Sections 4 and 5.

4. Research Gap and Contribution

The framework proposed in this study addresses the gaps identified in Section 3.4 through a unified design: an intent-aware AI system that detects clause-level risk, scores it using FMEA's Risk Priority Number model (Stamatis, 2003), aligns outputs with the full ISO 31000 governance lifecycle, and embeds Responsible AI safeguards calibrated to the healthcare regulatory environment. The central purpose is to equip

CLM teams with the analytical infrastructure needed to move from reactive dispute resolution toward systematic, proactive risk governance.

5. Methodology and Approach

This study adopts a Design Science Research (DSR) methodology (Peffers et al., 2007; Hevner et al., 2004), which focuses on designing and evaluating an artifact to solve a recognized practical problem. This paper covers through the demonstration and early evaluation phase; specific calibration parameters will be reported in the complete dissertation. Sections 5.1 through 5.6 describe the framework's core components: clause construction, intent interpretation, FMEA scoring, ISO 31000 mapping, Responsible AI safeguards, and the integrated pipeline.

5.1 Clause Scenario Construction

The pilot clause dataset was built in two phases. First, clauses from established legal NLP benchmarks—LEDGAR and CUAD—were mapped to three healthcare contract risk categories (fee schedule and reimbursement, prior authorization, and denials and appeals). Second, domain-specific clauses were authored by practitioners familiar with payer–provider agreements and CMS model contracts, then refined using large language models for linguistic diversity. Each clause passed through a multi-stage quality pipeline detailed in the dissertation.

5.2 Intent Interpretation and Ambiguity Detection

Intent interpretation uses a dual-model architecture. A fine-tuned Legal-BERT model (SLM) classifies each clause into one of three primary healthcare contract categories and produces a language clarity score. In parallel, GPT-4 (LLM) extracts structured intent tags from a predefined taxonomy, assesses scope breadth and operational triggers, and identifies payer tactics and protective language. These complementary outputs—deterministic classification from the SLM and semantic signals from the LLM—feed into a multi-layered guardrail architecture before risk scoring. The pipeline enforces a confidence threshold of ≥ 0.85, routing outputs into auto-accept, human-review, or reject bands. Each clause undergoes two-pass extraction; field-level agreement is measured via Jaccard similarity and Cohen's κ, with disagreements routed to human review. Cross-model validation compares LLM intent outputs against SLM classification to surface inconsistencies, and all extracted intents are validated against the predefined taxonomy. Clauses failing any validation layer are routed to human oversight rather than automated scoring. Threshold values and inter-rater reliability statistics will be detailed in the dissertation.

Ambiguity detection combines the SLM's clarity classification with the LLM's assessment of scope breadth, protective language (stabilizers), and conditional versus unconditional payer actions. When these signals diverge, the clause is flagged for human review rather than automated scoring. The divergence criteria are documented in the dissertation.

5.3 FMEA-Based Risk Scoring

Once SLM and LLM outputs are validated, the framework applies deterministic FMEA-based risk scoring. Zero LLM involvement occurs at this stage—the rule engine maps validated signals to scores through deterministic rules, ensuring reproducible, auditable results. Each clause is evaluated across three dimensions: Severity (potential impact, informed by clause type and intent tags), Occurrence (likelihood of triggering, informed by scope, payer tactics, and operational triggers), and Detection (difficulty of identifying the risk through manual review, informed by language clarity). All dimensions are scored on a 1–10 scale and combined into a weighted Risk Priority Number (RPN) that assigns each clause to a risk tier, with a severity override that escalates clauses scoring $S \geq 9$ directly to critical. For clauses with multiple risk signals, the engine generates a per-failure-mode decomposition, producing a separate FMEA row for each identified risk. Scoring calibration and threshold boundaries will be reported in the dissertation.

5.4 ISO 31000 Stage Mapping

The framework operationalizes ISO 31000's four lifecycle stages. Identify classifies clause text and extracts risk signals through the SLM–LLM pipeline with guardrail validation. Evaluate produces S, O, D scores and a composite RPN with full audit trace. Treat maps each risk tier to a governance strategy—Avoid, Reduce, Transfer, or

Accept—with healthcare-specific actions; high-severity clauses trigger mandatory human escalation. Monitor tracks clauses through configurable review cycles and automated re-evaluation triggers. Mapping rules are documented in the dissertation.

5.5 Responsible AI Safeguards

The framework embeds Responsible AI principles across the pipeline rather than as post-hoc additions. Explainability is achieved through reasoning traces documenting ambiguity, intent interpretation, and scoring rationale for each clause. Transparency is maintained through end-to-end audit trails—every decision is timestamped and machine-readable from classification through treatment. Bias control is enforced through auditing of scoring distributions across clause types and payer categories. Privacy is preserved through PHI exclusion at the data ingestion layer. Human oversight checkpoints at both the guardrail boundary and governance output stage ensure high-stakes decisions are never fully automated.

5.6 Framework Pipeline

Figure 1 illustrates the full operational pipeline. Clause input enters through ingestion where text is segmented and categorized. Signal extraction identifies clause type and intent. Outputs pass through guardrail validation; low-confidence results are routed to human oversight. Validated outputs proceed to FMEA-based risk scoring, producing a risk tier. The ISO 31000 mapping layer routes each scored clause to governance action across Identify, Evaluate, Treat, and Monitor. The final output is a governance record comprising a risk register entry, a recommended action, and a traceable audit trail.

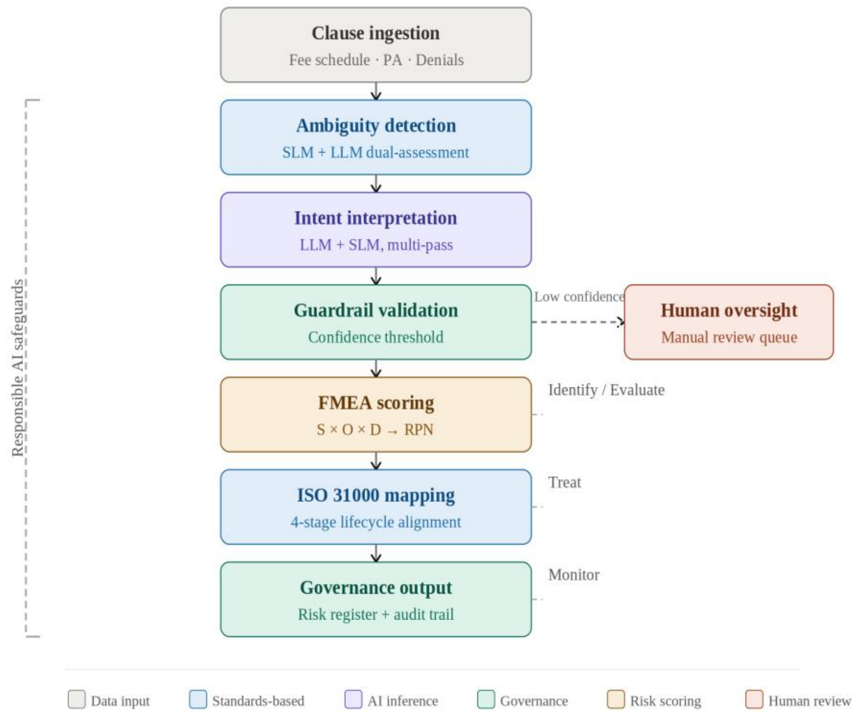

Figure 1. AI-driven clause-level risk governance pipeline

Table 2

Pipeline Walkthrough — Illustrative Clause Example

Stage	**Activity**	**Illustrative Output**
Input	Clause text ingested	*"Provider shall be reimbursed at 110% of the Medicare fee schedule. Payer reserves the right to adjust reimbursement rates annually based on market conditions, with thirty (30) days prior written notice. Any overpayments identified through audit may be recouped from future payments without prior authorization from Provider."*
Identify	SLM classification + LLM intent extraction	Clause type: fee_schedule (high confidence); Intents: financial risk exposure,

		payer-initiated recoupment, unilateral rate adjustment; Ambiguity flagged: “reserves the right” (unilateral action language), “market conditions” (undefined criteria); Model confidence: auto-accepted (≥ 0.85 threshold met); Cross-model validation: SLM–LLM agreement confirmed
Evaluate	FMEA scoring (S × O × D)	Severity: elevated — direct revenue impact with unilateral recoupment rights; Occurrence: moderate — audit-triggered recoupment is a documented industry practice; Detection: elevated — recoupment rights embedded within reimbursement terms, increasing manual review difficulty; Composite risk tier: Medium
Treat	ISO 31000 treatment selection	Strategy: Reduce — define objective criteria for “market conditions” adjustments, add dispute resolution period before recoupment, establish scope and lookback limitations on audit rights; Escalation: flagged for senior contract reviewer due to overlapping risk intents
Monitor	Regular monitoring and lifecycle tracking	Clause registered for quarterly review (medium-tier frequency); automated re-evaluation triggered if CMS updates fee schedule rules or payer issues rate adjustment notice; recoupment activity tracked against audit trail for drift detection

Note. Specific scoring values, thresholds, and calibration parameters are reserved for the complete dissertation.

6. Pilot Illustration and Early Findings

Table 3 presents three representative clause examples from the pilot phase to illustrate how the framework operates in practice. Each example covers one of the three primary clause categories, the identified risk intent, an illustrative risk tier, and the resulting ISO 31000 governance action. Risk tiers are presented qualitatively to demonstrate the scoring logic; specific calibration values will be reported exclusively in the dissertation. Clause examples are constructed from regulatory sources and do not contain data from any specific contract.

Table 3

Illustrative Pilot Examples: Clause Scenarios and Governance Outputs

Clause Excerpt (Representative)	Identified Risk Intent	Risk Tier (Illustrative)	ISO 31000 Stage & Governance Action
"Prior authorization is required for all elective procedures." (No response timeline specified.)	Undefined payer response timeline; exposure to indefinite payment deferral.	**High**	Evaluate → Treat: Flag for contract amendment; require insertion of payer response deadline; escalate to compliance review if unresolved.
"Fee schedule rates are subject to annual adjustment at the payer's	Unilateral rate adjustment without notice; revenue uncertainty for provider planning.	**Moderate–High**	Evaluate → Monitor: Document exposure; recommend minimum advance written notice

discretion." (No notification requirement specified.)			clause; monitor at contract renewal.
"Appeals must be submitted within 30 days of the denial notice." (Calendar vs. business days unspecified.)	Ambiguous day-count definition; risk of forfeiting appeal rights through misinterpretation.	**Moderate–High**	Treat → Monitor: Require explicit clarification of day-count definition; implement internal tracking alert set ahead of the deadline.

Note. Risk tiers are illustrative and intended to demonstrate the scoring logic. Specific calibration parameters will be reported exclusively in the complete dissertation.

The examples illustrate a key feature of the framework: governance actions are calibrated to the character and reversibility of the risk, not merely its numeric score. The denial-appeal clause, though rated moderate-high alongside the fee schedule scenario, warrants immediate treatment because misinterpretation permanently extinguishes the provider’s legal remedy—a consequence that is irreversible rather than merely costly. Early pilot observations also suggest the framework surfaces ambiguity patterns that standard classification-based approaches do not detect, connecting each finding to a specific, auditable governance action.

7. Reflection: Challenges and Learning

Three challenges have shaped the development of this research. First, data availability is limited. Clause-level datasets linking healthcare contract provisions to financial or regulatory outcomes are not publicly available. While resources such as CUAD and LEDGAR provide clause annotations, they do not capture operational consequences such as denials, underpayments, or audit triggers. Representative scenarios were therefore constructed using regulatory guidance, documented disputes, and domain expertise, as described in Section 5.1.

Second, explainability in regulated environments places demands on output design that go beyond technical accuracy. Risk assessments must be transparent enough that contract managers and compliance officers can interrogate, challenge, and defend them—requirements that proved difficult to satisfy through machine-generated outputs alone and necessitated iterative refinement of the framework's record structure. Third, the framework navigates an inherent tension between domain specificity and generalizability. Deep calibration to healthcare payer–provider contracting improves analytical precision but narrows the framework's immediate transferability to other industries. The current design optimizes for the healthcare regulatory context, with generalization treated as a deliberate objective for subsequent research phases.

The primary insight from this development phase is that the framework's value lies not only in technical risk detection but also in creating a structured language for risk governance that CLM teams can apply without requiring deep AI expertise.

8. Planned Next Steps

The immediate next step is expert validation of the scoring logic and intent interpretation approach, conducted with healthcare contracting practitioners and risk management professionals whose domain knowledge can stress-test the framework's assumptions. Following validation, the framework will be applied to a curated set of anonymized clause scenarios, with outputs benchmarked against documented dispute outcomes where these are available. The longer-term ambition is to produce a fully validated framework specification and actionable design recommendations for integrating clause-level risk intelligence into existing CLM platforms—contributing both to academic knowledge and to practitioner capability in healthcare contract governance.

9. Ethical and Regulatory Considerations

This study does not involve human participants. Clause examples used in the pilot phase are either synthetically generated or derived from publicly documented disputes and regulatory materials. Any future use of real contract language will follow applicable confidentiality, HIPAA, and institutional review requirements. The Responsible AI safeguards embedded throughout the framework are described in detail in Section 5.5.

References

Al-Mhdawi, M. K. S., Qazi, A., Alzarrad, A., Dacre, N., Rahimian, F., Buniya, M. K., & Zhang, H. (2023). Expert evaluation of ChatGPT performance for risk management process based on ISO 31000 standard. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4504409

American Hospital Association. (2023). Costs of caring report. https://www.aha.org/costsofcaring

Bizzaro, P. G., Della Valentina, E., Napolitano, M., Mana, N., & Zancanaro, M. (2024). Annotation and classification of relevant clauses in terms-and-conditions contracts. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). https://aclanthology.org/2024.lrec-main.108.pdf

Black Book Market Research. (2019). Improving contract management for healthcare. https://www.contractlogix.com/contract-management/improving-contract-management-for-healthcare

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). Legal-BERT: The muppets straight out of law school. In Findings of EMNLP 2020 (pp. 2898–2904). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.261

Chhatwal, R., Gronvall, P., Huber-Fliflet, N., Keeling, R., Zhang, J., & Zhao, H. (2019). Explainable text classification in legal document review [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1904.01721

Hendrycks, D., Burns, C., Chen, A., & Ball, S. (2021). CUAD: An expert-annotated NLP dataset for legal contract review. In Proceedings of the Neural Information Processing

Systems Track on Datasets and Benchmarks (NeurIPS 2021). https://doi.org/10.48550/arXiv.2103.06268

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105. https://doi.org/10.2307/25148625

International Organization for Standardization. (2018). ISO 31000:2018 Risk management – Guidelines. https://www.iso.org/standard/65694.html

Ittan, A. (2024). AI-powered contract risk scoring: The development of a contract risk scoring engine. International Journal of Research in Computer Applications and Information Technology, 7(2), 2643–2653. https://doi.org/10.34218/IJRCAIT_07_02_200

Koreeda, Y., & Manning, C. D. (2021). ContractNLI: A dataset for document-level natural language inference for contracts. Findings of EMNLP. https://doi.org/10.48550/arXiv.2110.01799

Leivaditi, S., Rossi, J., & Kanoulas, E. (2020). A benchmark for lease contract review. arXiv. https://doi.org/10.48550/arXiv.2010.10386

Lippi, M., Pałka, P., Contissa, G., Lagioia, F., Micklitz, H.-W., Sartor, G., & Torroni, P. (2019). CLAUDETTE: An automated detector of potentially unfair clauses in online terms of service. Artificial Intelligence and Law, 27, 117–139. https://doi.org/10.1007/s10506-019-09243-2

Mahoney, C., Gronvall, P., Huber-Fliflet, N., & Zhang, J. (2022). Explainable text classification techniques in legal document review. 2022 IEEE International Conference on Big Data (pp. 2044–2051). https://doi.org/10.1109/bigdata55660.2022.10020626

Moon, S., Chi, S., & Im, S. (2022). Automated detection of contractual risk clauses from construction specifications using BERT. Automation in Construction, 140, 104465. https://doi.org/10.1016/j.autcon.2022.104465

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of

Management Information Systems, 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Raptopoulos, P., Filandrianos, G., Lymperaiou, M., & Stamou, G. (2025). PAKTON: A multi-agent framework for question answering in long legal agreements. arXiv. https://doi.org/10.48550/arXiv.2506.00608

Ruggeri, F., Lagioia, F., Lippi, M., & Torroni, P. (2021). Detecting and explaining unfairness in consumer contracts through memory networks. Artificial Intelligence and Law, 30(1), 59–92. https://doi.org/10.1007/s10506-021-09288-2

Sader, S., Husti, I., & Daroczi, M. (2020). Enhancing failure mode and effects analysis using auto machine learning: A case study of the agricultural machinery industry. Processes, 8(2), 224. https://doi.org/10.3390/pr8020224

Sonani, R., Govindarajan, V., & Verma, P. (2025). Federated learning-driven privacy-preserving framework for decentralized data analysis and anomaly detection in contract review. International Journal of Advanced Computer Science and Applications, 16(3). https://doi.org/10.14569/ijacsa.2025.0160301

Song, W., Rong, W., & Tang, Y. (2024a). Quantifying risk of service failure in customer complaints: A textual analysis–based FMEA approach integrating LDA, sentiment polarity, and entropy with CRITIC and TOPSIS. Advanced Engineering Informatics, 60, 102377. https://doi.org/10.1016/j.aei.2024.102377

Song, W., & Zheng, J. (2024). A new approach to risk assessment in failure mode and effects analysis based on engineering textual data. Quality Engineering. https://doi.org/10.1080/08982112.2024.2304815

Stamatis, D. H. (2003). Failure mode and effect analysis: FMEA from theory to execution (2nd ed.). ASQ Quality Press.

Tuggener, D., von Däniken, P., Peetz, T., & Cieliebak, M. (2020). LEDGAR: A large-scale multi-label corpus for text classification of legal provisions in contracts. In Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020). https://aclanthology.org/2020.lrec-1.155.pdf

Wong, S., Zheng, C., Su, X., & Tang, Y. (2023). Construction contract risk identification based on knowledge-augmented language model (arXiv preprint arXiv:2309.12626). arXiv. https://doi.org/10.48550/arXiv.2309.12626

World Commerce & Contracting. (2024a). The purpose of contracts: How clarity of purpose is required to measure contract quality and effectiveness. WorldCC. https://www.worldcc.com/Portals/IACCM/Reports/Purpose-of-Contracts.pdf

Xu, W., Deng, Y., Lei, W., Zhao, W., Chua, T.-S., & Lam, W. (2022). ConReader: Exploring implicit relations in contracts for contract clause extraction. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2581–2594). https://aclanthology.org/2022.emnlp-main.166.pdf