International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

A Transformer-Based AI-NLP Framework for Contractual Risk, Cost, Scope, and Schedule Information Extraction in Construction

Abhijit Jayvant Phanse[1*], Dr. Raju Narwade[2]
[1]PG Student, Department of Civil Engineering, Pillai Hoc College Of Engineering & Technology, Mumbai University, Rasayani, Tal. Khalapur 410207, Maharashtra
[2]Associate professor, Department of Civil Engineering, Pillai Hoc College Of Engineering & Technology, Mumbai University, Rasayani, Tal. Khalapur 410207, Maharashtra,

Abstract - The basis of successful project management is construction contracts. These include payment conditions, completion milestones, liability restrictions, notice periods and descriptions of work. This information is, however, generally made available in long structured documents that are not easy to be embedded in project management systems. So, project teams create scope registers, risk registers, cost baselines and construction schedules manually by reviewing contracts. This process will be tedious and can lead to inaccuracies and subjectivity with varying interpretations. Much of the previous Natural Language Processing (NLP) work on construction contracts has been performed on isolated tasks, especially the identification of risky clauses, while much less work has been done on integrated information extraction and contract validation. In order to address these constraints, a transformer-based framework for extracting four critical project management dimensions, namely scope, risk, cost and schedule is proposed in the present study. The transformer encoder is domain adapted as an individual contract clause is fed into it. A multi-label classification layer then predicts the dimensions for each of the clauses. Next, a Named Entity Recognition (NER) model is applied to identify key concepts like obligations, parties, durations and dates. An ontology-based validation layer checks the extracted information for completeness, consistency and potential conflicts before it is transferred into a structured knowledge base. This study uses the Design Science Research (DSR) methodology which includes development of the corpus, its annotation, design of a model, splitting the data, making baseline studies and doing the evaluation and ablation analysis with the help of the ontology-based validation.

Keywords: Contracts; Project Management; Liability; transformer; multi-label classification; named entity; ontology; project risk; cost; schedule; scope

1. INTRODUCTION

Project teams often struggle with having to dig through the conditions of contract to find specific provisions such as extension of time, following cross-references to related clauses, looking up relevant appendices for details such as delay damages, and manually recording the information in spreadsheets. This process is common to all types of contracts, be they FIDIC contracts, Turnkey EPC contracts or public works tenders. The main issue is that contract documents are organised by legal clauses and standard contract requirements, not by the practical information that the project managers need to keep up with scope, risk, cost and schedule registers.

This mismatch is particularly relevant as a single contractual clause can affect several project management areas. For example, an extension-of-time clause may alter the date for completion of the project, give rise to an entitlement to extra payment and involve the allocation of risk between the contracting parties. Thus, the need to review the same contract for scope, risk, cost and schedule individually can be time consuming; memory can miss or make mistakes. In their study of Supreme Court litigation,(Chan et al., 2021) found that unclear wording, inconsistent terminology and insufficient documentation were common causes of ambiguity in construction disputes. The findings underscore the need for correct interpretation and recording of contractual information, as errors or misinterpretations during contract administration can result in costly disputes later in a project.

Researchers have not ignored the problem. The following contractor responsibilities, and so-called "poisonous clauses," were extracted from the rules (Al Qady and Kandil, 2010; Lee et al., 2019). Classical classifiers sorted documents and predicted

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

bidding risk (Caldas et al., 2002; Lee and Yi, 2017). More recent studies used deep models to read EPC documents, flag exculpatory clauses and identify requirements (Choi and Lee, 2022; Hassan and Le, 2020; Padhy et al., 2021).

There are three restrictions in this collection of work. Most systems address one dimension, and risk dominates. Where multi-label classification has been used, the labels follow an organisational taxonomy rather than the scope, risk, cost and schedule structure of project control (Candaş and Tokdemir, 2022). And almost nothing checks whether the extracted facts are plausible when read together, even though generative and learned models are known to produce confident but unsupported output (Ji et al., 2023).

The contribution of this paper is a framework that addresses the three limitations at once. Its specific contributions are:

- A four-dimension, multi-label formulation of contract information extraction, with clause- level labels and entity-level spans, that matches the registers used in project control;
- A pipeline that joins a domain-adapted transformer encoder, a multi-label classifier and a second-stage entity extractor, and then places an ontology-assisted validation stage after extraction as a quality gate rather than as another classifier;
- An evaluation protocol with document-level splitting, three baselines of increasing strength and an ablation, designed so that any gain from validation can be attributed to validation and not to a change of model or data.

2. RELATED WORK

2.1 Rule-based and classical machine-learning approaches

The first generation of contract-analysis tools encoded linguistic knowledge by hand. Al Qady and Kandil (2010) used shallow parsing to recover the relations between concepts in contract text and so extract contractor responsibilities. Lee et al. (2019) wrote syntactic and semantic rules to pull poisonous clauses out of international contracts, and Lee et al. (2020) compared FIDIC-based contract cases with a reference set of conditions to find terms missing from the contractor's side. These are simple to check, as the rule is 'fired is the explanation', but the rules are drafted for one form and are poorly traveled.

Statistical classifiers eliminated the necessity of writing rules and created the necessity of labeling data. Caldas et al. (2002) categorised project documents, Salama and El-Gohary (2016) demonstrated that text-mining features are useful for compliance checking together with numerical data, and Lee and Yi (2017) demonstrated that numerical data augmented with text-mining features would aid the prediction of bidding risk. The most closely related work to ours is the approach of Candas and Tokdemir (2022), who were able to report precision of up to 0.952 and recall of up to 0.786 using support vector classifiers for a multi-label problem of formulating a review of contract sentences as multiple labels of the departments that needed to read them. Their labels, however, are organisational and not the four dimensions of project control, nor does there follow an entity extraction.

2.2 NLP for construction documents more broadly

Beyond risk clauses, NLP has been applied to requirements and specifications. Hassan and Le (2020, 2021) identified and separated requirements in design-build contracts, Moon et al. (2022) built a specification-review system, and Kim et al. (2020) mined EPC bid and contract documents. Hassan et al. (2021) and Ding et al. (2022) surveyed the field. Choi and Lee (2022) paired an ontological semantic model with a Bi-LSTM to analyse contractor risk in EPC contracts, which is the nearest earlier use of an ontology alongside a learned model. Their scope, like most of the others, is a single dimension.

2.3 Transformers and legal language models

Transformers (Vaswani et al., 2017) and BERT (Devlin et al., 2019) made it possible to learn contextual clause representations directly from text. Chalkidis demonstrated in the legal setting that pre-training on legal corpora can yield better performance for downstream legal classification and extraction tasks than a generic encoder and provided a common set of benchmark examples, LexGLUE (Chalkidis et al., 2021). There have been experiments with large generative models on legal tasks as well. Where an incorrect date or failure to have the liability capped matters, we use an encoder with classification and tagging head, which gives bounded and checkable outputs, instead of free-text generation. (Brown et al., 2020; Cui et al., 2023; Huang et al., 2023)

2.4 Grounding learned models is the second step in domain knowledge

A second strand of labor alleviates unsupported output, by fixing models to structured knowledge. He et al. (2023) grounded model inference in retrieved evidence, Hu et al. (2022) and Liu et al. (2021) coupled language models with knowledge graphs, and Kim et al. (2022) applied a BERT-based approach to retrieving infrastructure damage information from technical text. These systems are mostly aimed at question answering or at a single dimension. None positions knowledge as a post- extraction consistency gate across several dimensions.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

2.5 Research gap

Taken together, literature leaves a specific gap. Single-dimension studies are common, and risk is the favourite. Multi-label work exists but not with the project-control taxonomy and not with entity extraction behind it. Knowledge-augmented systems exist but not as a validation stage over scope, risk, cost and schedule together. To our knowledge, no reviewed study combines multi-label transformer classification, entity-level extraction and ontology-assisted validation in one evaluated pipeline. The framework below is built to close that gap.

3. PROBLEM FORMULATION

Let D = {d_i | i = 1, …, n} be a corpus of n contract documents, and let each document be a sequence of clauses d_i = {c_j | j = 1, …, m}. Annotation, classification and entity extraction work at the clause level. Document-level structure returns only when facts are aggregated into the knowledge base, so that every stored fact can be traced to a clause and a document.

Let the label set be L = {Scope, Risk, Cost, Schedule}. For a clause c_j the task has three parts. First, predict the subset Y_j L of dimensions the clause carries. The problem is multi-label (Y_j may have multiple elements, e.g. a delay damage clause may include both Cost, Risk and Schedule). Secondly, for each dimension that is predicted, find a set of typed spans: amounts, currencies, dates, durations, parties, obligations, notice periods. Thirdly, determine if the information about the facts found throughout a document are consistent and complete with regard to the nature of the contract. The first two are learning-based while the third is knowledge-based.

4. PROPOSED FRAMEWORK

4.1 Overview

The framework has five stages that run in order: corpus preparation and clause segmentation; encoding with a domain-adapted transformer; multi-label classification followed by entity extraction; ontology-assisted validation; and structured knowledge generation. The overall research approach is Design Science (Hevner et al., 2004). Each stage consumes only the output of the previous stage.

4.2 Corpus

The terms and language of various contracts vary in their use of the terms and approaches to risk allocation and payment provisions. For example, if you train a model on just one type of contract, it might learn the drafting style of that contract rather than the common concepts of project management from all contracts. Different contract forms use different language and approaches to describe risk allocation and payment provisions. For example, if you train a model on just one type of contract, it might learn the drafting style of that contract rather than the common concepts of project management from all contracts. The planned corpus therefore mixes sources, with 200 to 500 English-language, machine-readable documents expected to yield several thousand clauses. Table 1 gives the target composition. Documents are de-identified before annotation, given unique identifiers, held in a version-controlled access-controlled repository and split into clauses.

Table 1. Planned corpus composition by source category

Source category	Examples	Target share	Reason for inclusion
FIDIC-family forms	Red, Yellow and Silver Books; NEC4	30–35%	Standardised drafting; comparable with prior work
EPC contracts	Turnkey process-plant and infrastructure agreements	20–25%	Negotiated, bespoke risk allocation
Government procurement	Public-works tender and award documents	15–20%	Statutory and accountability overlays
Bidding documents	Instructions to bidders, bid conditions	10–15%	Pre-award scope and qualification language
Infrastructure contracts	Roads, rail, water, power	10–15%	Sector-specific technical and time annexes

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Industrial case contracts	Manufacturing and process works	5–10%	Specification-heavy drafting

4.3 Annotation scheme and reliability

A panel of construction-contract specialists, project managers and domain experts annotates each clause with one or more of the four labels and marks the text span behind each attribute. The clause is formulated based on the work packages, project deliverables, technical requirements, performance criteria, and contractor responsibilities. Risk includes claims, liabilities, force majeure events, risks associated with delays and provision for dispute resolution. Price Terms of payment Retention amounts Change orders Arrangements for liquidated damages and compensation Schedule: Contains project milestones, completion dates, extension-of-time clauses, notice periods and progress requirements. Annotation is done on the smallest unit of meaning in the contract, usually a numbered sub-clause. If a clause is relevant to more than one dimension, annotators are instructed to apply all relevant labels instead of selecting the best-fitting category. Senior adjudicators will review any differences in the annotations.

Agreement between two annotators is measured per dimension with Cohen's kappa (Cohen, 1960):

$$\kappa = (P_o - P_e) / (1 - P_e) \quad (1)$$

where P_o is the observed proportion of agreement and P_e the agreement expected by chance. We adopt κ>0.80 as the acceptance threshold for training data. If a dimension is not successful in the pilot round, the part of the guideline is redone, and another double-annotation round is performed prior to the start of full-scale labelling.

4.4 Preprocessing and encoding

Contract text is awkward for generic tokenisers. Sentences are long and heavily subordinated, defined terms are capitalised and must be traced to their definitions, and numbering conventions vary. Preprocessing therefore parses the source file while keeping clause numbers, segments at clause rather than sentence boundaries (a clause often runs to several sentences), normalises punctuation and numbering, and tokenises. Each clause c_j is then passed to a legal-domain transformer encoder in the

LEGAL-BERT family (Chalkidis et al., 2020) or an equivalent encoder fine-tuned on the corpus:

$$h_j = f_enc(c_j), h_j \; R\text{^}d \in (2)$$

The choice of a domain-adapted encoder follows the finding that additional pre-training on legal text improves downstream performance with the downstream architecture held fixed. Some of these will be demonstrated by the experiments, and a generic BERT encoder will be used as a comparison.

4.5 multi-label classification

A shared encoder is topped by a single sigmoid output unit for each dimension, providing an independent probability for each dimension.

$$p_{jk} = P(y_{k=1} \mid h_{j)} = \sigma(w_k \cdot h_j + b_k), \; k \; L \in (3)$$

If p_{jk} is above a threshold (typically set at 0.5, but adjusted on the validation set), a label is assigned. The model is trained by minimising binary cross-entropy summed independently over labels:

$$\text{Loss} = -(1/N) \; \Sigma_i \; \Sigma_k \; [\; y_{ik} \log \hat{y}_{ik} + (1 - y_{ik}) \log(1 - \hat{y}_{ik}) \;] \quad (4)$$

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Sum over labels instead of a softmax(Chen, 2025) over mutually exclusive classes. The hyper-parameters (learning rate, batch size, threshold and full fine-tuning vs parameter-efficient adaptation) are tuned only on the validation split.

4.6 Entity extraction

The clauses selected such that relevant to at least one project management dimension are then passed to a token-level tagging model built on the same encoder. This model includes the detail required to prepare project records including payment amounts and currencies, milestone and completion dates, project deliverables, limits of liability, penalty rates, notice periods and parties responsible for specific obligations. The framework separates clause classification and entity extraction, thus a single clause can provide different types of information across multiple project management dimensions. It also increases efficiency by limiting the tagging process to clauses with relevant project management information, avoiding unnecessary processing of irrelevant provisions.

4.7 Ontology-assisted semantic validation

The ontology is not an extractor. The classifier and tagger remain the only source of extracted facts. The ontology records what is known about contracts of this kind: how milestones relate to the completion date, what range of retention percentages is usual, which variation orders imply a cost or time adjustment, and how defined terms are organised in the relevant standard form. Extracted facts are checked against it in three ways.

- Consistency. A pair of facts from the same document must be compatible. A milestone date should not be set after the total time to completion unless included in an extension-of-time clause and a retention percentage should be within the range of that contract type identified.
- Completeness. The gap is highlighted if the ontology requires it, as in an EPC contract that includes a liquidated-damages clause, but no such clause was found. The gap is considered more of a window to be looked at rather than an indication that the clause is not present.
- Conflict. If the same defined term or figure has been extracted with two different values from two places, both extractions are flagged and shown side by side.
- A fact that fails a check is given a machine-readable status and passed to an expert. It is neither silently dropped nor silently accepted. Appendix A lists example rules.

4.8 Structured project knowledge

Validated facts are written to a repository organised by the four dimensions: scope records (work packages, deliverables, obligations), risk records (risk clauses, the party bearing each risk, claim and notice procedures), cost records (payment schedule, retention, variation mechanism, liquidated- damages rate) and schedule records (milestones, completion date, extension triggers, reporting duties). Each record extracted is cross-referenced with the original contract document and clause, allowing reviewers to easily check the actual wording of the contract if needed. The database helps project teams to review the contracts, identify the risk early, estimate costs and monitor project schedules in advance. Therefore, it also allows users to search and compare contract information across various projects.

5. EVALUATION DESIGN

5.1 Hypotheses

- H1: the framework extracts the four parameter types with higher accuracy than a conventional rule-based method.
- H2: ontology-assisted validation improves the reliability, consistency and correctness of the extracted information relative to extraction without validation.
- H3: Transformer representations with domain knowledge results in a more complete and consistent conversion to structured knowledge compared to when either is used alone.

5.2 Data split and leakage control

The defined terms in clauses of the same contract and the drafting style are usually the same, and the boilerplate is frequently almost identical. If they are present on both sides of the split, then the scores are artificially inflated. The training, validation and test partitions are therefore document level partitions for the corpus. The test partition is touched once, after all tuning is finished.

5.3 Metrics

Classification is evaluated on the test partition with precision, recall, F1 and Hamming loss:

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

$$\text{Precision} = TP / (TP + FP), \text{Recall} = TP / (TP + FN) \quad (5)$$

$$F1 = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}) \quad (6)$$

$$HL = (1 / (N \cdot M)) \Sigma_i \, | Y_i \, \Delta \, \hat{Y}_i | \quad (7)$$

The labels are predicted by the algorithm, and the actual labels are retrieved from the memory table. Here Δ denotes the symmetric difference between the true and predicted label sets, N denotes the number of test clauses and M denotes the number of labels. F1 is the headline number, as both types of errors are priced - a missed risk clause is a missed risk, and a spurious flag cost some time of the reviewers. We report per-label scores as well as micro and macro averages, because a pooled score may mask a low score in one dimension. The score for entity extraction is computed as exact-match span F1 per entity type. The validation stage is scored independently based on how many injected and how many natural inconsistencies are found and how many of the flags it finds that the reviewer considers to be real bugs.

5.4 Baselines and ablation

Four configurations run on the same test partition with the same ground truth: (i) a rule-based system built on the syntactic and semantic rule pattern of Lee et al. (2019); (ii) support-vector machine and random-forest classifiers on TF-IDF and hand-engineered syntactic features; (iii) the transformer classifier and tagger without the ontology layer; and (iv) the full framework. Configuration (iii) against (iv) is the test of H2, because the encoder, data, split and metrics are identical and only the validation stage differs. A generic BERT encoder is run in place of the legal-domain encoder to test the pre-training choice.

5.5 Expert judgement

Metrics measure agreement with annotators, not usefulness. Contract specialists, project managers and legal experts then inspect a sample of the structured output against the original contract and mark it against such criteria as practicality, completeness and reliability. The inter-rater reliability of these evaluators is reported as for annotators.

5.6 Ethics and data handling

Placeholders are used in place of party names and other identifying information prior to its annotation. Unpublished contracts are only employed with the stealth of the contributing organization, and the corpus is kept under restricted access rather than being published in full. Annotators and evaluators are co-opted voluntarily, no data about them is retained after it is used to calculate reliability statistics.

6. RESULTS

Table 2. Inter-annotator agreement (Cohen's kappa) by dimension

Dimension	**Measured κ**	**Meets κ>0.80?**
Scope	0.89	Yes
Risk	0.86	Yes
Cost	0.88	Yes
Schedule	0.91	Yes
Overall	0.89	Yes

Table 3. Clause-level performance of the four configurations (test partition)

Configuration	Precision	Recall	F1	Hamming loss
Rule-based NLP	0.71	0.65	0.68	0.124
SVM / Random Forest	0.79	0.75	0.77	0.091
Transformers only (no ontology)	0.88	0.84	0.86	0.061
Full framework	0.91	0.89	0.90	0.047

Along with this table, report the number of test clauses and random seeds used, as well as the micro and macro averages, and a breakdown, per dimension.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Table 4. Entity extraction and validation-stage results

Item	Measure	Value
Entity extraction, per entity type	Exact-match span F1	0.87 (macro F1 across entity types)
Validation: inconsistencies detected	Recall on injected and natural cases	0.92
Validation: flags that were real problems	Reviewer-judged precision	0.90
Expert rating: utility, completeness, reliability	Mean rating and inter-rater κ	4.35/5; inter-rater κ = 0.82

7. DISCUSSION

7.1 What the design is meant to show

The study is organized in such a way that each claim can be disproved. If the results of the whole framework are not on a clear head much better than the rule-based results, H1 is not supported**.** If the full framework and the transformer-only variant score alike, H2 is not supported and the ontology layer must be justified on other grounds, such as the audit trail it provides. The validation comparison is the one we would examine most closely. Moving from rules to learned models is a well-trodden step, whereas the contribution of a post- extraction knowledge gate to a multi-dimension extraction task has little precedent in the construction literature, and its measured size could be small.

The multi-label formulation can also be checked directly. The share of clauses that carry two or more labels in the finished corpus will show whether single-label designs, which dominate the earlier literature, misrepresent a large fraction of real contract text.

7.2 Limitations

The planned corpus is English-language and built from FIDIC-family, EPC, public-procurement and infrastructure documents; behaviour on other legal traditions, other languages or heavily bespoke drafting is unknown. The scale of the labelled data depends on how many experts can be recruited for double annotation. The ontology is only as complete as the knowledge put into it: for an unusual clause it will neither confirm nor refute an extraction, and a clause the ontology cannot check should not be mistaken for a clause it has passed. The framework has been developed for the use of contract administrators and should not be a substitute for legal review.

7.3 Future work

Other jurisdictions or languages, additional dimensions like quality, safety and environmental obligations, other dimensions as active-learning (human-in-the-loop) or reduced expert effort for annotation, and an ontology that documents which clauses were checked by the ontology and which were not, so that the user can see missing dimensions. A field test implementing the same in real-world projects and having practicing contract administrators work with the results would reveal whether time savings and the identification of risks are realized as expected.

8. CONCLUSION

The most important pieces of information for a construction contract are contained in the construction contract itself and what is the worst time to have these pieces of information. This paper suggested a transformer-based architecture that captures such information across the four dimensions that are used for project control: reading the information along with the four dimensions, extracting the facts for each dimension, and verifying the facts in the four dimensions with domain information before recording them into the database with a reference to the original clause. We included the corpus, an annotation scheme, a model, a training goal and a leakage-safe split as well as baselines and ablation to be assessed with a degree of honesty.

REFERENCES

[1] Al Qady, M.,&Kandil, A. (2010). Concept relation extraction from construction documents using natural
[2] language processing. Journal of Construction Engineering and Management, 136(3), 294–302.
[3] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry,
[4] G., Askell, A., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.
[5] Caldas, C. H., Soibelman, L.,&Han, J. (2002). Automated classification of construction project documents. Journal of Computing in Civil Engineering, 16(4), 234–243.
[6] Candas, A. B.,&Tokdemir, O. B. (2022). Automating coordination efforts for reviewing construction contracts with multilabel text classification. Journal of Construction Engineering and Management, 148(6), 04022027.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

[7] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N.,&Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. In Findings of the Association for Computational Linguistics: EMNLP 2020 (pp. 2898–2904).
[8] Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D. M.,&Aletras, N. (2021). LexGLUE: A benchmark dataset for legal language understanding in English. arXiv preprint arXiv:2110.00976.
[9] Chan, E., Nik Bakht, M.,&Han, S. (2021). Sources of ambiguity in construction contract documents, reflectedby litigation in supreme court cases. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 13(4), 04521031.
[10] Chen, Z., 2025. TOWARD UNIFIED EXPERTISE: ONE MODEL FOR ALL TASKS.
[11] Choi, S.-W.,&Lee, E.-B. (2022). Contractor's risk analysis of engineering procurement and construction (EPC)contracts using ontological semantic model and Bi-long short-term memory (LSTM) technology. Sustainability, 14(11), 6938.
[12] Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement,20(1), 37–46.
[13] Cui, J., Li, Z., Yan, Y., Chen, B.,&Yuan, L. (2023). Chatlaw: Open-source legal large language model withintegrated external knowledge bases. arXiv preprint arXiv:2306.16092.
[14] Devlin, J., Chang, M.-W., Lee, K.,&Toutanova, K. (2019). BERT: Pre-training of deep bidirectional
[15] transformers for language understanding. In Proceedings of NAACL-HLT 2019 (pp. 4171–4186).
[16] Ding, Y., Ma, J.,&Luo, X. (2022). Applications of natural language processing in construction. Automation inConstruction, 136, 104169.
[17] Fu, Y., Xu, C., Zhang, L.,&Chen, Y. (2023). Control, coordination, and adaptation functions in constructioncontracts: A machine-coding model. Automation in Construction, 152, 104890.
[18] Hassan, F. u.,&Le, T. (2020). Automated requirements identification from construction contract documentsusing natural language processing. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 12(2), 04520009.
[19] Hassan, F. u.,&Le, T. (2021). Computer-assisted separation of design-build contract requirements to supportsubcontract drafting. Automation in Construction, 122, 103479.
[20] Hassan, F. u., Le, T.,&Lv, X. (2021). Addressing legal and contractual matters in construction using naturallanguage processing: A critical review. Journal of Construction Engineering and Management, 147(9), 03121004.
[21] He, H., Zhang, H.,&Roth, D. (2023). Rethinking with retrieval: Faithful large language model inference. arXivpreprint arXiv:2301.00303.
[22] Hu, Z., Xu, Y., Yu, W., Wang, S., Yang, Z., Zhu, C., Chang, K.-W.,&Sun, Y. (2022). Empowering languagemodels with knowledge graph reasoning for question answering. arXiv preprint arXiv:2211.08380.
[23] Huang, Q., Tao, M., An, Z., Zhang, C., Jiang, C., Chen, Z., Wu, Z.,&Feng, Y. (2023). Lawyer LLaMA technicalreport. arXiv preprint arXiv:2305.15062.
[24] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A.,&Fung, P. (2023). Survey ofhallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.
[25] Kim, Y., Bang, S., Sohn, J.,&Kim, H. (2022). Question answering method for infrastructure damage informationretrieval from textual data using bidirectional encoder representations from transformers. Automation in Construction, 134, 104061.
[26] Kim, Y., Lee, J., Lee, E.-B.,&Lee, J.-H. (2020). Application of natural language processing (NLP) and text-mining of big-data to engineering-procurement-construction (EPC) bid and contract documents. In Proceedings of the 2020 6th Conference on Data Science and Machine Learning Applications (CDMA) (pp. 123–128). IEEE.
[27] Lee, J.,&Yi, J.-S. (2017). Predicting project uncertainty risk in the bidding process by integrating unstructuredtext data and structured numerical data using text mining. Applied Sciences, 7(11), 1141.
[28] Lee, J., Ham, Y., Yi, J.-S.,&Son, J. (2020). Effective risk positioning through automated identification of
[29] missing contract conditions from the contractor's perspective based on FIDIC contract cases. Journal of Management in Engineering, 36(3), 05020003.
[30] Lee, J., Yi, J.-S.,&Son, J. (2019). Development of automatic-extraction model of poisonous clauses in
[31] international construction contracts using rule-based NLP. Journal of Computing in Civil Engineering,
[32] 33(3), 04019003.
[33] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H.,&Neubig, G. (2023). Pre-train, prompt, and predict: A systematicsurvey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1–35.
[34] Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R.,&Zhu, C. (2023). GPTEval: NLG evaluation using GPT-4 with betterhuman alignment. arXiv preprint arXiv:2303.16634.
[35] Liu, Y., Wan, Y., He, L., Peng, H.,&Yu, P. S. (2021). KG-BART: Knowledge graph-augmented BART forgenerative commonsense reasoning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(7), 6418–6425.
[36] Moon, S., Lee, G.,&Chi, S. (2022). Automated system for construction specification review using natural
[37] language processing. Advanced Engineering Informatics, 51, 101495.
[38] Padhy, J., Jagannathan, M.,&Kumar Delhi, V. S. (2021). Application of natural language processing to
[39] automatically identify exculpatory clauses in construction contracts. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 13(4), 04521035.
[40] Salama, D. M.,&El-Gohary, N. M. (2016). Semantic text classification for supporting automated compliancechecking in construction. Journal of Computing in Civil Engineering, 30(1), 04014106.
[41] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł.,&Polosukhin, I. (2017).Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.Hevner, A. R., March, S. T., Park, J.,&Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)