Automated Construction Contract Summarization Using Natural Language Processing and Deep Learning

X. Xue[1], Y. Hou[1], and J. Zhang[1*]

[1] School of Construction Management Technology, Purdue University, USA
E-mail: xue39@purdue.edu, hou144@purdue.edu, (*corresponding author) zhan3062@purdue.edu

Abstract –

The interpretation of construction contracts is crucial to the management and success of a project. Correct and accurate interpretation could support the smooth construction of high-quality built assets. Misunderstanding and omissions may lead to costly rework and delay. One main challenge of construction contract interpretation lies in the length of construction contracts. Therefore, the demand for reducing the length of such documents while keeping their main information elements emerges. To address this research need, the authors proposed to use natural language processing (NLP) and deep learning technology to summarize construction contracts (i.e., text summarization). There are many deep learning models available and developed for text summarization. However, their performance on construction contracts is to be tested. To address this gap, the authors proposed a new merit-based evaluation method to evaluate the performance of three deep learning models on text summarization of construction contracts, which were reported the state-of-the-art performance on text summarization tasks in general English corpus. The proposed method evaluated selected models from three aspects: information completeness, information correctness, and human readability. The Distilbart model, which scored 5.23, 4.82, and 5.05 in these three aspects, respectively, outperformed the other models in all three aspects.

Keywords –

Deep Learning; Automated Construction Contract Summarization; Text Summarization; Natural Language Processing; Artificial Intelligence; Construction Management

1 Introduction

Construction contract is a critical type of construction document that details the terms agreed upon by all involved stakeholders [1]. In general, construction contracts specify critical provisions such as payment schedules, construction costs, and completion dates. Additionally, construction contracts also specify how disputes should be resolved when raised, and other procedural agreements. To ensure that all necessary information is included, and ambiguity is avoided, building contracts typically try to cover every aspect that might potentially be predicted [2]. Therefore, construction contracts may easily become too long for human readers to digest easily.

The length of construction contracts places a high cognitive burden on their human readers and increases the time required to understand and process the contractual information. As a result, reducing the length of such contracts while maintaining their main idea emerges as an urgent research need. The main idea that a body of text conveys can be split into many information elements. Text summarization is the process of creating a condensed version of the body of text by retaining critical information elements and removing uncritical ones [3]. In recent years, many transformer-based deep learning models reported the state-of-the-art performance on the task of text summarization. However, because they were mostly tested on datasets of general English corpus, their performance on domain-specific texts such as construction contracts is not clear.

Evaluating the performance of text summarization is a challenging task [4]. The majority of text summarization metrics were concerned with determining the similarity between automatically generated summaries and some target summaries [5]. These metrics assume that the greater the similarity is, the better the summarizations are and, consequently, the better the model performance is. The target summaries were typically generated by experts and demonstrated what a good summary of the entire text should look like. One benefit of using such metrics is that the measurement can be fully automated. Their shortcomings, on the other hand, are also significant. Certain automated metrics, for example, precision, recall, and f1-score at the word level, do not take into account the sequential order of words in the summaries [5]. The same set of words arranged in a different sequence may have a significantly different meaning or (in the extreme case) make no sense at all.

There were also metrics that take word sequence into account, such as the BiLingual Evaluation Understudy (BLEU) [6] and the Metric for Evaluation of Translation with Explicit ORdering (METEOR) [7]. But they are known to award summaries that make little/no sense or are difficult to read for humans.

To address the above-mentioned problem, the authors proposed a new merit-based text summarization evaluation method that focuses on human perception of the summarization results. Three transformer-based deep learning models were then evaluated using the proposed method in this research: (1) Distilbart [8], (2) Pegasus [9], and (3) Bidirectional and Auto-Regressive Transformer (BART) [10]. These evaluated models were selected because they have been reported with the state-of-the-art performance on many different natural language processing (NLP) tasks, such as machine translation, question and answering, text summarization, and named entity recognition (NER). However, their performance on the summarization of construction contracts has not been tested to the best of the authors' knowledge. This research provides an initial evaluation of the performance of transformer models on the summarization of construction contracts.

2 Background

2.1 Natural Language Processing

Natural language processing (NLP) has a wide range of applications in the architecture, engineering, and construction (AEC) domain. Research on NLP in the AEC domain mostly focused on classifying or extracting information from construction documents (e.g., construction contracts, and building codes) for future processing by human or machine [11]. For example, Caldas and Soibelman [32] used machine learning and NLP in the classification of construction management documents. Zhang and El-Gohary [12] combined sematic NLP-based information extraction with automated reasoning to accomplish automated reasoning with building code requirements to support compliance checking. Xue et al. [13] proposed a semi-automated method to extract regulatory information from tables in building codes. Zhang and El-Gohary [14] developed a deep learning-based information extraction system for extracting regulatory requirements to support automated building code compliance checking. Xue and Zhang [15] increased the accuracy of part-of-speech tagging of building codes by an error-fixing method. Li and Cai [16] utilized NLP in the processing of infrastructure requirements to support compliance checking of underground utility lines. Le and Jeong [17] used NLP techniques, such as Word2Vec, to classify semantic relation between terminologies in transportation asset manuals. Dimyadi et al. [33] developed a table-based NLP algorithm to convert building codes from normative text to computable rules. Song et al. [34] used deep learning method to convert Korea building codes to predict-argument structure. Al-Qady and Kandil [36] leveraged sematic parsing to extract relations between concepts in construction contraction clauses.

2.2 Text Summarization

The main goal of text summarization is to preserve the main information elements of a body of text while reducing its length [3]. Text summarization systems allow users to obtain the main information elements of documents without having to read the full text. In general, there are single-document systems [18] that summarize one document at a time, and multi-document systems that summarize multiple documents into one summary [19]. Based on the approach of summarization, text summarization systems can be classified into extractive systems that extract important sentences from the documents [20], and abstractive systems that aim to generate a summary by reducing unimportant information from the original documents by processing computerized representations of text (e.g., embedding vector) [21].

Literature review suggested that summarization of contractual clauses is challenging because of the difference between legal document and general document and the limited amount of available training dataset. For example, Manor and Li [37] prepared a dataset of contract documents and evaluated existing text summarization methods on it. They concluded that the summarization of contract documents is challenging because the gap between legal documents and training texts of existing summarization methods is substantial. Elnaggar et al. [38] launched a dataset of legal documents for the tasks of translation, summarization, and classification. They utilized transfer learning (leveraging pre-trained models) and multi-task learning (jointly training one model in multiple tasks) to reach the state-of-the art performance on these three tasks on a dataset of a few thousand sentences. However, their research cannot be applied directly to this study because their research focused on legal documents in Europe, which is drastically different from construction contracts in the United States.

2.3 Transformer Models

Transformer models belong to one type of deep learning model that frequently achieved the state-of-the-art performances in various NLP tasks in recent years [22]. The core of a transformer model is the attention mechanism, which transforms one sequence to another sequence using an encoder-decoder structure [23]. The recurrent neural network (RNN) is known not to perform well at capturing long-term dependency. The attention mechanism effectively solved this shortcoming by learning the importance of each token and/or correlations between each token in the source sequence and each token in the target sequence during the training phase. However, the computational complexity of calculating every possible pair of tokens in each sequence is high. Local attention mechanisms (i.e., a subtype of attention mechanism) emerged as a remedy by only calculating attention between surrounding words [24, 25]. Self-attention is another subtype of attention mechanism where the source sequence and the target sequence are the same [26]. Self-attention has been proven to be effective in tasks such as machine translation, named entity recognition, and text summarization. Transformer models are usually pre-trained on self-supervised tasks, such as predicting masked tokens or predicting the next token given all the previous tokens in a sentence [27, 28].

Transformer models are constructed in an encoder-decoder fashion. The encoder receives the input sequence (i.e., text) in the form of embedding vectors. The encoder converted the input vectors to a new vector called "internal state" after processing them. The decoder is then informed of the internal state and generates the outputs (i.e., summary). Typically, transformers are trained on semi-supervised auto-regressive tasks, such as predicting the next word given previous words and predicting masked words given their surrounding words. However, the detail of each transformer varies. BART's encoder was trained to predict masked words, whereas its decoder was trained to predict next word given previous words [10]. Pegasus was trained to generate gap sentences between sentences [9]. The distillbart model is nearly identical to the BART model in most aspects except that it is pre-trained using knowledge distillation, which is a technique that reduces the size of deep learning models by using a larger deep learning model to train a smaller deep learning model. While the larger deep learning model is trained to generate predictions, the smaller deep learning model is trained to match the predictions of the larger model [8]. An early form of knowledge distillation uses data labeled by the larger model (instead of human annotators) to train the smaller model. Knowledge distillation in its current form focuses on matching the behavior of the output layer of two models. The smaller model is trained jointly to make the same predictions and assign the same probability of all predicted classes as those by the larger model [48].

2.4 Challenges

The evaluation of text summarization has been a challenge. Most automated evaluation methods would compare the summarization of a specific model to a gold standard of summarization and generate a score based on predefined metrics. The more similar the machine-generated summarization is to the gold standard, the higher the score will be. This type of metrics is based on two underlying assumptions both of which may not be very robust. First, it assumes text summarization has a ground truth, and the closer a summarization is to the ground truth, the better the summarization is. However, a text may have multiple good summaries based on different information organizations and expressions. In addition, it is hard for machines to measure how human readers would understand a summarization. Second, it assumes automated text summarization evaluation means n-gram similarity between machine-generated summarization and the gold standard. While the problem of the first assumption may be solved by generating multiple gold standards and averaging scores from all gold standard versions, the cost of such evaluation could increase significantly. The second assumption is also questionable because the meaning of a sentence is very sensitive to the sequence of words, and high word/token-level similarity does not guarantee similar meaning. A slight shift in the sequence of words may completely change the meaning of a sentence.

3 Methodology

It is common to pre-process textual data before it was fed into deep learning models. To analyze the impact of pre-processing on the summarization of construction contracts, the authors compared performance of deep learning models without and with pre-processing. In the first setting, the textual data was fed into deep learning models directly without pre-processing. The inputs to deep learning models are therefore unaltered construction contract text. In the second setting, inputs went through the following pre-processing steps: (1) tokenization (i.e., break down strings of sentences into lists of words and punctuations), (2) lowercasing (i.e., convert all characters into lower case), (3) stop word removal (remove common words that do not carry context-specific information, such as "a" "an" "the"), (4) removal of number, punctuation, and underscores (i.e., fill-in-the-blank space from construction contract template), and (5) lemmatization (reduce inflectional forms of words to their base forms).

The authors proposed a merit-based evaluation of the performance of the state-of-the-art deep learning transformer models on the summarization of construction contracts in this research. The merit-based evaluation has three steps. First, construction contract clauses that are suitable for the summarization text are collected. Clauses that do not need to be summarized (too short) or cannot be summarized properly (contain too much blank space and miss too much information) are removed in the step. In the second step, transformer models generate summaries of the selected clauses. In the third step, human reviewers manually evaluate the transformer models by evaluating the summaries generated by the models.

Manual evaluation of transformer models was used to avoid the shortfalls associated with automated evaluation. Manual evaluation by domain experts, as an alternative to automated evaluation, can have the following two advantages. First, manual evaluation ensures that a summarization with high score is indeed easy to understand for human readers. Second, manual evaluations can effectively assess the meanings carried by the summarizations. The proposed evaluation quantifies transformer performance on three dimensions: information completeness, information correctness, and human readability. Information completeness is a metric that measures if a summary retains the key information elements from the original text. Information correctness is a metric that measures how accurate a summary is. A good summary should accurately represent the original full text. Human readability refers to the ease with which the summary can be understood. Models' performance in each category is rated on a scale of 0 to 10. The higher the score, the better the performance is. Domain experts are allowed to rate performance according to their understanding of construction contract based on their construction domain knowledge. They are also required to maintain a high-level objectivity and consistency across the evaluation.

4 Experiment

For the purpose of this research, the authors collected publicly available and free-to-download construction contracts or contract templates online. In total, nine construction contracts or contract templates were collected [31, 39-46]. The collected construction contracts were cleaned to fit the needs of the research. Some paragraphs were removed because they were either too short or contained too much blank space. A dataset that contains ninety-two paragraphs of construction contracts or contract templates was prepared. After the dataset was created, a careful evaluation and extensive literature review were conducted to select transformer candidates for the evaluation. Three transformers were selected: (1) Distilbart [8], (2) Pegasus [9], and (3) BART [10] (Table 1). The selected models all have achieved the state-of-the-art performance or near state-of-the-art performance on the task of text summarization in large open datasets of general English corpus [8, 9, 10]. Because transformer models were usually published in different configurations, one configuration of each model was selected based on the popularity of the configuration. For Distilbart, the selected configuration was distilbart-cnn-12-6. For BART, the selected configuration was bart-large-cnn. For Pegasus, the selected configuration was pegasus-xsum. These configurations were meaningful expressions. For example, the "cnn" in the distilbart-cnn-12-6 means convolutional neural network is used. The 12 refers to 12 layers of encoders. The "large" in bart-large-cnn means this configuration has a large number of parameters. The "xsum" in the pegasus-xsum means the configuration was trained on the xsum dataset [29]. For each model, the configuration with the largest number of downloads in Huggingface Transformer model hub [30] was selected. The selected models weren't altered or fine-tuned in any way for the evaluation.

Table 1. Selected Models

Model Name	Configuration
BART	distilbart-cnn-12-6
Pegasus	pegasus-xsum
Distilbart	bart-large-cnn

Each selected transformer model generated a summary of each paragraph in the dataset with and without preprocessing of the textual data. The summaries were then evaluated by the authors using the merit-based evaluated method. All models were run on a desktop computer with an RTX 3090 graphic card and a Ryzen 3950x CPU.

5 Result

The performance of transformer models with and without pre-processing was summarized in Tables 2, 3 and 4.

Table 2. Model Performance with Pre-processing

	BART	Pegasus	Distilbart	Average
Completeness	4.62	2.12	4.70	3.81
Correctness	4.59	2.16	4.57	3.77
Human Readability	4.62	2.51	4.45	3.86
Overall	4.61	2.26	4.57	3.81

Table 3. Model Performance without Pre-processing

	BART	Pegasus	Distilbart	Average
Completeness	4.74	4.98	5.23	4.98
Correctness	4.72	4.6	4.82	4.71
Human Readability	4.66	4.82	5.05	4.84
Overall	4.71	4.8	5.03	4.76

Table 4. Model Performance Comparison

	With Pre-processing	Without Pre-processing
Completeness	3.81	4.98
Correctness	3.77	4.71
Human Readability	3.86	4.84
Overall	3.81	4.76

The performance of deep learning models was better when inputs were not pre-processed rather than when they were pre-processed. When the inputs were not pre-processed, the Distilbart model outperformed the other two transformer models in all three aspects. This shows the Distilbart model could be a good choice for the summarization of construction contracts. For future refinement on the text summarization of construction contracts, the Distilbart model should be a good starting point. It is very likely that the fine tuning of the model will lead to better performance, which is suggested in literature [38]. Example summarizations by each model are provided in Table 5.

Table 5: Example Summary without Pre-Processing

Deep Learning Model	Original Text	Summary
BART	"Owner shall have a right to inspect the Work at any time and request that Contractor promptly correct any Work that is defective or does not conform to the Contract Documents. If required, the Work shall be inspected and certified by the appropriate state or local agency or health officer at each necessary stage." [31]	Owner shall have a right to inspect the Work at any time and request that Contractor promptly correct any Work that is defective
Pegasus		The owner of the property where the Work is to be performed has the right to inspect the Work at any time.
Distilbart		Owner shall have a right to inspect the Work at any time and request that Contractor promptly correct any Work that is defective

The performance was not optimal, as no model achieved a score greater than six. The Distilbart model had the highest score among the three, making it a good choice for starting contract summarization for now. However, it is strongly recommended that a model dedicated to the construction contract text summarization task be developed (i.e., trained on construction contracts corpus). The task of fine-tuning transformer models has been identified as a high-priority area of research. Among all evaluation criteria, the information correctness criterion had the lowest average, indicating that increasing the information accuracy of the summarization is a pressing need.

6 Discussion

The Distilbart model outperformed the other two models in the evaluation. The authors attributed its success to its training strategy and the process of knowledge distillation. The BART model employed a decoder that had been trained to generate the next word given previous words in a sentence, implying that it is capable of generating new texts (summary). Its encoder is trained to predict masked words, enhancing the model's robustness. The Distilbart model inherited the BART model's advantages. Additionally, the knowledge distillation process enhanced the model's generalizability (i.e., ability to generate good summary).

The gap in the performance of deep learning models with and without pre-processing could be attributed to two reasons. First, deep learning models had difficulty in generating complete sentences. It is likely that stop words, although did not carry a lot of useful information themselves, were necessary for the transformer models to generate complete sentences. Second, the Pegasus model generated a lot of summaries, such as, "All photographs courtesy of AFP, EPA, Getty Images, and Reuters" and "All photographs are copyrighted." that had no correlation with the core of the original text. Therefore, more comprehensive evaluation is necessary before putting deep learning models into practice in construction contract text summarization. The behaviours of deep learning models currently are not guaranteed to perform as users may expect. Based on our experimental results, the performance of deep learning model decreased when the construction contract input is pre-processed. Therefore, to increase the accuracy of summarization, pre-processing that was typically used in other NLP tasks may need to be avoided.

7 Contributions to the Body of Knowledge

This research makes three distinct contributions to the body of knowledge. First, the authors proposed an evaluation method for text summarization based on merit in three different dimensions: information completeness, information correctness, and human readability. The proposed method ensures that the summary's evaluated score corresponds to its human perception. By involving humans in the evaluation process, the method avoids awarding high scores to summaries with poor human perception. Second, a recommended deep learning model was identified for contract summarization by comparing the performance of three different but popular models: BART, Pegasus, and Distilbart. The Distilbart model outperformed other models on all three criteria, making it a good starting point for the task of summarizing construction contracts. Last but not least, it was found that in contrast to typical NLP tasks, pre-processing does not necessarily help in the construction contract summarization using deep learning models.

8 Limitations and Future Work

The following limitations are acknowledged. First, the evaluation is composed of three components. While these provided a relatively holistic evaluation comparing to the state of the art, additional factors can be included to provide a more detailed and comprehensive evaluation, such as, grammar, structure, and coherence [35]. Second, the proposed evaluation method necessitates manual efforts, which could be costly and time-consuming. Future research could investigate how to make the evaluation more efficient. In addition, human subjectivity may influence the outcome of an evaluation. Certain factors, such as personal preference and educational background, may influence how certain summaries are rated. Future research should investigate how to minimize the impact of human subjectivity on such evaluation.

9 Conclusion

Existing research that leverages NLP in the construction domain mostly directly adopted benchmarks and metrics from other domains, the suitability of which is seldom asked. In this paper, the authors proposed a new merit-based evaluation method for text summarization. The proposed method was used in evaluating the performance of three deep learning transformer models on the text summarization of construction contracts or contract templates. The evaluation was conducted manually by rating performance from three aspects: information completeness, information correctness, and human readability. Three models were selected based on their popularity: Distilbart, Pegasus, and BART. The Distilbart model outperformed the other two models in all three aspects. The performance of deep learning modes was found to be better when inputs were not pre-processed than when the inputs were pre-processed. As the direct application of such deep learning model on the domain specific construction contracts did not achieve very high scores, the authors suggest that the fine-tuning of Distilbart model and/or retraining it based on domain corpus should be in the future direction of text summarization of construction contracts.

10 References

[1] Chan E. E. Nik-Bakht M. and Han S. H. Sources of ambiguity in construction contract documents, reflected by litigation in supreme court cases. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 13(4): 04521031, 2021.

[2] Bagherian-Marandi N. and Akbarzadeh M.R. Two-layered fuzzy logic-based model for predicting court decisions in construction contract disputes. *Artificial Intelligence and Law*, 29(4): 453-484, 2021.

[3] Abualigah L. Bashabsheh M. Alabool Q. H. and Shehab M. Text summarization: a brief review. *Recent Advances in NLP: The Case of Arabic Language*, 1-15, 2020.

[4] Gambhir M. and Gupta V. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1): 1-66, 2017.

[5] Ermakova L. Cossu J. V. and Mothe J. A survey on evaluation of summarization methods. *Information Processing & Management*, 56(5): 1794-1814, 2019.

[6] Papineni K. Roukos S. Ward T. and Zhu W.J. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311-318, Philadelphia, Pennsylvania, 2002.

[7] Banerjee S. and Lavie A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, 2005, pp. 65-72.

[8] Shleifer S. and Rush A. M. Pre-trained summarization distillation. *arXiv preprint.* arXiv:2010.13002, 2020.

[9] Zhang J. Zhao Y. Saleh M. and Liu P. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *arXiv preprint.* arXiv: 1912.08777, 2020.

[10] Lewis M. Liu Y. Goyal N. Ghazvininejad M. Mohamed A. Levy O. Stoyanov V. and Zettlemoyer L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. a*rXiv preprint*. arXiv:1910.13461, 2019.

[11] Wu C. Li X. Guo Y. Wang J. Ren Z. Wang M. and Yang Z. Natural language processing for smart construction: current status and future directions. *Automation in Construction*, 134:104059, 2022.

[12] Zhang J. and El-Gohary N. M. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. *Automation in Construction*, 73:45-57, 2017.

[13] Xue X. Wu J. and Zhang J. Semiautomated generation of logic rules for tabular information in building codes to support automated code compliance checking. *Journal of Computing in Civil Engineering*, 36(1):04021033, 2022.

[14] Zhang R. and El-Gohary N. A deep neural network-based method for deep information extraction using transfer learning strategies to support automated compliance checking. *Automation in Construction*, 132:103834, 2021.

[15] Xue X. and Zhang J. Building codes part-of-speech tagging performance improvement by error-driven transformational rules. *Journal of Computing in Civil Engineering*, 34(5): 04020035, 2020.

[16] Li S. Cai H. and Kamat Vineet R. Integrating natural language processing and spatial reasoning for utility compliance checking. *Journal of Construction Engineering and Management*, 142(12):04016074, 2016.

[17] Le T. and Jeong H. NLP-based approach to semantic classification of heterogeneous transportation asset data terminology. *Journal of Computing in Civil Engineering*, 31(6):04017057, 2017.

[18] Christian H. Agus M. P. and Suhartono D. Single document automatic text summarization using term frequency-inverse document fequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, 7(4):285-294, 2016.

[19] Uçkan T. and Karcı A. Extractive multi-document text summarization based on graph independent sets. *Egyptian Informatics Journal*, 21(3):145-157, 2020.

[20] Ferreira R. Cabral L. Lins R. Silva G. Fred Freitas Cavalcanti G. Lima R. Simske S. and Favaro L. Assessing sentence scoring techniques for extractive text summarization. *Expert systems with applications*, 40(14):5755-5764, 2013.

[21] Song S. Huang H. and Ruan T. Abstractive text summarization using LSTM-CNN based deep learning. *Multimedia Tools and Applications*, 78(1): 857-875, 2019.

[22] Kalyan K. S. Rajasekharan A. and Sangeetha S. Ammus: A survey of transformer-based pretrained models in natural language processing. *arXiv preprint,* arXiv:2108.05542, 2021.

[23] Vaswani A. Shazeer N. Parmar N. Uszkoreit J. Jones L. Gomez A. N. Kaiser L. and Polosukhin I. Attention is all you need. In *Proceedings of Advances in Neural Information Processing Systems*, pages. 5998-6008, Long Beach, California, 2017.

[24] Tjandra A. Sakti S. and Nakamura S. Local Monotonic attention mechanism for end-To-end speech and language processing. *arXiv preprint,* arXiv:1705.08091, 2017.

[25] Li L. Tang S. Deng L. Zhang Y. and Tian Q. Image Caption with global-local attention. In *Proceedings of the Thirty-first Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence*, pages 4133-4139, San Francisco, California, 2017.

[26] Shaw P. Uszkoreit J. and Vaswani A. Self-attention with relative position representations. *arXiv preprint*, arXiv:1803.02155, 2018.

[27] Devlin J. Chang M. W. Lee K. and Toutanova K. Bert: pre-training of deep bidirectional transformers for language Understanding. *arXiv preprint,* arXiv:1810.04805, 2018.

[28] Brown T. B. Mann B. Ryder N. Subbiah M. Kaplan J. Dhariwal P. Neelakantan A. Shyam P. Sastry G. Askell A. Agarwal S. Herbert-Voss A. Krueger G. Henighan T. Child R. Ramesh A. Ziegler D. M. Wu J. Winter C. Hesse C. Chen M. Sigler E. Litwin M. Gray S. Chess B. Clark J. Berner C. McCandlish S. Radford A. Sutskever I. and Amodei D. Language models are few-shot learners. *arXiv preprint*, arXiv:2005.14165, 2020.

[29] Narayan S. Cohen S. B. and Lapata M. Don't give me the details, just the summary! Topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages,1797-1807, Brussels, Belgiuml, 2018.

[30] Huggingface.co. Model hub. On-line: https://huggingface.co/models, Accessed: 07/01/2022.

[31] County of Montrose. Construction contract. On-line: https://www.montrosecounty.net/DocumentCenter/View/823/Sample-Construction-Contract, Accessed: 17/01/2022.

[32] Caldas C.H. and Soibelman L. Automating hierarchical document classification for construction management information systems. *Automation in Construction*, 12(4), 395-406, 2003.

[33] Dimyadi J. Fernando S. Davies K. and Amor R. Computerising the New Zealand building code for automated compliance audit. In *Proceeding of New Zealand Built Environment Research Symposium,* pages 39-46, Auckland, New Zealand, 2020.

[34] Song J. Lee J.K. Choi J. and Kim I. Deep learning-based extraction of Predicate-Argument Structure

(PAS) in building design rule sentences. *Journal of Computational Design and Engineering*, 7(5):563-576, 2020.

[35] Allahyari M. Pouriyeh S. Assefi M. Safaei S. Trippe, E.D. Gutierrez J.B. and Kochut K. Text summarization techniques: a brief survey. *arXiv preprint,* arXiv:1707.02268, 2017.

[36] Al-Qady. M. and Kandil. A. Concept relation extraction from construction documents using natural language processing. *Journal of Construction Engineering and Management,* 136(3): 294-302, 2017.

[37] Manor. L. and Li. J. J. Plain English summarization of contracts. *arXiv preprint,* arXiv:1906.00424, 2019.

[38] Elnaggar A. Gebendorfer C. Glaser I. Matthes F. Multi-task deep learning for legal document translation, summarization and multi-label classification. In *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*, pages 9-15, New York, New York, 2018.

[39] Zhuang. F. Qi. Z. Duan. K. Xi. D. Zhu. Y. Zhu. H. Xiong. H. and He. Q. A comprehensive survey on transfer learning. In *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE)*, pages43-76, Piscataway, New Jersey, 2020.

[40] Legaltemplates. The party hiring the general contractor to complete construction on his or her property. what is a construction agreement. On-line: https://legaltemplates.net/form/lt/construction-contract-agreement, Accessed:08/02/2022.

[41] Eforms. Construction contract template. On-line: https://eforms.com/employment/independent-contractor/construction, Accessed:08/02/2022

[42] Signaturely. Construction contract agreement template. On-line: https://signaturely.com/contracts/construction-contract-agreement-template, Accessed: 08/02/2022.

[43] BuildingAdvisor. Model construction contract for homeowners. On-line: https://buildingadvisor.com/project-management/contracts/model-construction-contract_1, Accessed: 08/02/2022.

[44] Washington LLC. Imperium Grays Harbor LLC Construction Agreement Cost Plus Fee. On-line: https://www.sec.gov/Archives/edgar/data/1381697/000119312507121562/dex107.htm, Accessed: 08/02/2022.

[45] City of Colorado Springs. Sample contract template. On-line: https://coloradosprings.gov/sites/default/files/construction_sample_contract_template_version_100316.pdf, Accessed: 08/02/2022.

[46] Florida Contractors. The construction agreement. On-line: http://floridacontractors.net/downloads/Residential%20Contract%201.pdf, Accessed: 08/02/2022.

[47] The Board of County Commissioners of Escambia County, Florida. Standard construction contract documents. On-line: https://myescambia.com/sites/myescambia.com/files/A_AE.pdf, Accessed: 08/02/2022.

[48] Hinton G. Vinyals O. and Dean J. Distilling the knowledge in a neural network. *arXiv preprint. arXiv:1503.02531* 2.7 (2015).