UNB EST. 1785 UNIVERSITY OF NEW BRUNSWICK | Off-site Construction Research Centre

Northumbria University NEWCASTLE

Transforming Construction with Off-site Methods and Technologies (TCOT) Conference: Designing Tomorrow's Construction, Today

August 20-22, 2024, Fredericton, New Brunswick, Canada

Review of Natural Language Processing (NLP) And Generative AI Applications in the Era of Construction 4.0

Mozaffari, M. Hamed[1,*], Popov, Pavel[1], RazaviAlavi, SeyedReza[1], and Jalaei, Farzad[1]
[1] Construction Research Centre, National Research Council Canada, Government of Canada, Ottawa, ON K1A 0R6, Canada
[*] mhamed.mozaffarimaaref@nrc-cnrc.gc.ca

Abstract: Artificial Intelligence plays a pivotal role in the era of Construction 4.0 to promote construction automation and digitalization. The construction industry produces a significant amount of textual data, most of which is unstructured and stored in various formats, such as construction specifications, reports, drawings, and contracts. The increasing volume of unstructured textual data requires the development of big data analytical and automatic information retrieval tools. Natural language processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence that enables the automatic analysis and representation of human language. Recently, there have been significant breakthroughs in the literature on NLP and AI, especially with the introduction of state-of-the-art generative AI techniques such as large language models (LLMs). These techniques reduce manual intervention, accelerate processes, handle large volumes of data, and are adaptable to a wide range of applications. This paper aims to review and categorize the existing literature on the use of NLP and Generative AI techniques in building information modelling (BIM) data interoperability, compliance checks with building codes, and information management. We will analyze trends, identify research gaps, and evaluate the benefits and limitations of each method for the construction industry in the era of Construction 4.0.

Keywords: Natural Language Processing; Construction 4.0; Building Information Modelling; Generative AI; Large Language Models; NLP for Construction

1 INTRODUCTION

Industry 4.0 significantly improves the way of manufacturing through ubiquitous connectivity and informed decision-making enabled by big data and artificial intelligence (AI). In the context of Industry 4.0, the construction field, including architecture, engineering, and construction (AEC), named Construction 4.0 aiming to develop in the direction of digitalization and intelligence to achieve significant improvements in automation, productivity, and reliability (Ding, Ma, and Luo 2022). Construction 4.0 heavily relies on data (structured and unstructured data) to build and maintain the interaction between the physical and virtual worlds. A significant amount of data generated by stakeholders of construction projects is textual information. Of these textual data, 80% are unstructured (Shamshiri, Ryu, and Park 2024; Chung et al. 2023; Wu et al. 2022; Peng et al. 2023), and they are stored in various formats, such as e-mails, construction specifications, non-conformance reports, accident reports, facility inspection reports, drawings, and contracts (Chung et al. 2023; Ding, Ma, and Luo 2022) in different storage file extensions, such as doc files, extensible markup language (XML), hypertext markup language (HTML), portable document format (PDF), computer-aided design (CAD and RVT), industrial foundation classes (IFC), etc.

Natural language processing (NLP) and text mining (TM) are a branch of artificial intelligence that deals with interactions between humans and computers. TM extract information from unstructured and structured text data with no semantic considerations, while NLP benefits from machine learning techniques that can process, understand, and simulate human language abilities (Shamshiri et al.,2024). NLP provides an intelligent way to process text data, enabling the intelligent agent to learn from human language and automatically complete knowledge representation (KRep), retrieval, and reasoning processes in a human-like way (Ding, Ma, and Luo 2022). The overall trend of literature on NLP and TM in construction has been reviewed by (Shamshiri, Ryu, and Park 2024). This review paper aims to review and categorize the existing literature on the use of NLP and Generative AI techniques in building information modelling (BIM) data interoperability, compliance checks with building codes, and information management.

We collected key publications in the literature from Scopus and Google Scholar using pre-defined terms in article titles, abstracts, and keywords, such as “natural language processing”, “project management”, “automated construction”, “laws and legislation”, “industry foundation classes”, “Construction 4.0”, “Generative AI”, “automatic compliance check”, “information management”, “data interoperability”, and more. Our initial search from 1980 to 2024 returned back 4215 articles. Narrowing down article selection by combining keywords and filtering alternative names (e.g., “Building Information Model” vs. “BIM”) and non-relevant articles, we got the result of 357 articles. We filtered out our search more by limiting it to the articles with the main emphasis on the use of NLP or Generative AI for digital construction, and we ended up with 48 articles. An interesting observation was that the final selected articles were published from 2010 to 2024, which has a correlation with the publication analysis of other review papers in the field (Shamshiri, Ryu, and Park 2024; Ding, Ma, and Luo 2022). Figure 1 illustrates a network clustering result of the main keywords in our review articles. As can be seen from the network, Generative AI has not been consolidated yet in the literature while NLP is relatively investigated in various domains. From the figure, it is also clear that NLP has a strong connection to automated compliance checking, information management, and knowledge management.

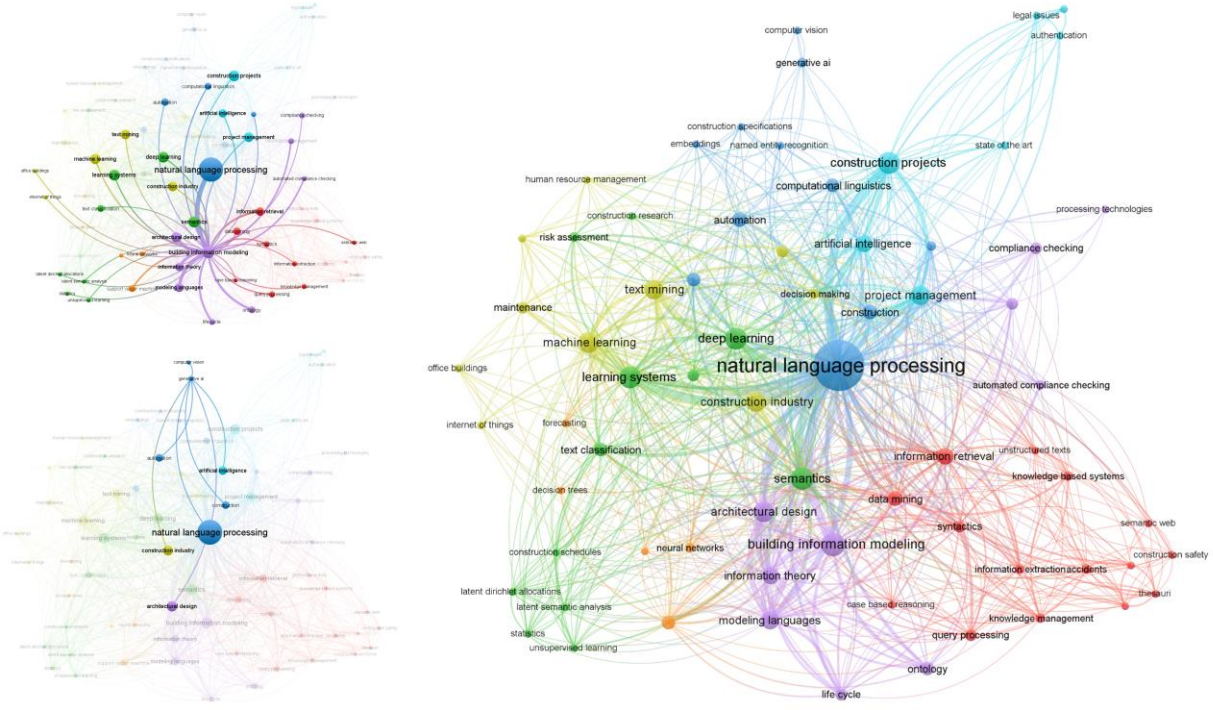

Figure 1. Network visualization and clustering of keyword occurrences regarding NLP and Generative AI for Construction.

2 Natural Language Processing and Generative AI

NLP involves two major steps, syntax analysis and semantic analysis, to convert human language into a format that computers can read. Syntax analysis consists of morpheme analysis, part-of-speech (POS) tagging and parsing, whereas the meaning of words and sentences is analyzed based on grammatical relationships between words. Conventional NLP was based on human-defined rules, which made it hard for computers to understand language semantics. WordNet (Kilgarriff and Fellbaum 2000) is an English lexical database that provides sets of synonyms that are linked through semantic relations. Machine learning and deep learning have changed the methodology of the NLP process and have improved accuracy dramatically. By deploying an extensive amount of data and neural network algorithms, computers can learn semantics in natural language by themselves. Word embedding is one of the major techniques for the semantic analysis of words, where a computer learns the meaning of words from distributed representation in vector space. In general, the evolution of NLP can be divided into three main eras from 1970 to 2010 (Locatelli et al. 2021): 1) rule-based systems based on complex sets of manually written rules, 2) statistical inference systems based on statistical models, and 3) deep learning approaches based on neural network algorithms. After these era, pre-trained large language models are the common trend and method of use in the literature, such as the Bidirectional Encoder Representations from Transformers (BERT) and OpenAI's GPT models based on the concept of transformers. Figure 2 shows important terms defined with their connections to NLP, Linguistics, and Human Language.

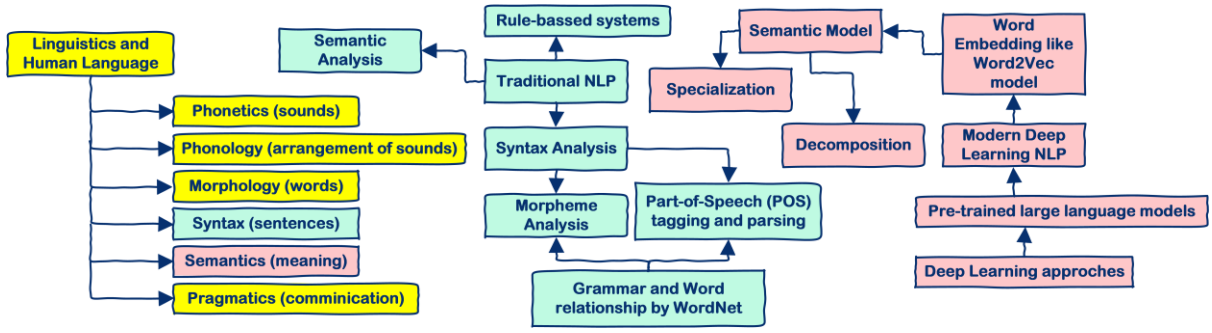

Figure 2. A few important terms in NLP and Linguistics and their connections are extracted from the NLP and linguistics literature.

NLP has also been utilized for digital transformation and smart construction when information is extracted from unstructured data, including from building information modelling (BIM) or digitalized text data, construction project management when knowledge discovery technologies contribute to the automated development of relations between concepts, known as ontology, contract risk analysis by providing automated compliance checking and identifying risky clauses from contractual documents, such as specifications, and facility management benefits by extracting cause-effect analyses of infrastructure from facility inspection reports and mining user experiences from complaint data (Chung et al. 2023). The primary tasks of NLP can generally be divided into natural language understanding (NLU) and natural language generation (NLG) (Ding et al., 2022).

3 Related Reviews

Numerous review papers exist in the literature. (Chung et al. 2023) systematically attempted to compare state-of-the-art NLP techniques in the computer science domain and their applications in the construction domain. This review work provides an invaluable list of influential papers on state-of-the-art NLP in computer science. (Di Giuda, Locatelli, and Seghezzi 2020; Yu Wang et al. 2020) reviewed recent applications of NLP in construction up to the year 2019. they found that NLP in construction is used for information management, procurement management, risk management, automated compliance checking (ACC), construction management and safety. The categorization method of (Wu et al. 2022) recognized

five core steps to apply NLP for smart construction: data collection, pre-processing, feature engineering, text analyzing and understanding, and performance evaluation. They concluded that text data on social media is very large, while much less data is collected for project-level applications. (Wu et al. 2022) provided an excellent brief review of NLP methods (e.g., supervised and unsupervised methods), processing stages (e.g., tokenization, vectorization, etc.), evaluation criteria and other NLP methods (e.g., distance metrics, rule base techniques, association, etc.), applications of NLP in smart construction with detail examples (e.g., extraction of data for risk prediction), and a framework on NLP in smart construction.

Construction projects often encounter legal issues such as claims, disputes, and litigation, which can lead to delays, increased costs, and poor relationships between contractors. (F. ul Hassan, Le, and Lv 2021; F. U. Hassan and Le 2022) conduct a literature review on contractual data needs, and they classified NLP methods for legal document applications. Another review study was completed by (Erfani and Cui 2021), and they enumerated usage cases of NLP for construction contracts, design requirements, risk registers, change orders, claims and litigation documents, and safety reports. (Hong et al. 2021) attempted to review and compare NLP-based clustering methods for construction schedules clustering. (Locatelli et al. 2021) accomplished a quantitative review to explore the state-of-the-art NLP methods in the AEC sector up to the year 2021, especially for BIM.

Generally, the deep learning-based NLP methods used in AEC include the two tasks of text classification (TC), which divides texts into different groups and Named Entity Recognition (NER) in the AEC domain, which detects semantic elements in a sentence. (Zheng et al. 2022) developed a large-scale domain corpus and a benchmark for pre-trained domain-specific deep-learning language models for the AEC domain. (Erfani and Cui 2021) indicated that the power of the novel deep learning-based approach, such as the BERT model, lies in its ability to capture contextual meaning. In another review work for contract information retrieval, (Pham and Han 2023) classified NLP-based publications from 2011 to 2021 based on algorithm and task. They also conducted case studies using the BERT model in comparison with other machine learning models for contract risk identification, allocation, and response. Their work showed that BERT could perform with higher accuracy compared to other machine learning models, such as bi-directional long-short-term memory (BiLSTM).

Recently, the emergence and rapid adoption of advanced large language models (LLMs) like OpenAI's GPT, Google's PaLM, and Meta's Llama have shown great potential and sparked considerable global interest (Ghimire, Kim, and Acharya 2024). However, despite rapid advancements in AI transforming many industry practices, construction largely lags in this adoption. (Ghimire, Kim, and Acharya 2024) conducted a literature review on what will be the future use of Generative AI in the construction industry and what are opportunities and challenges. Another literature review on generative AI for building structural design was conducted by (Liao et al. 2024). Both studies showed the potential of Generative AI in construction and focused more on design and data exchange. To mention a few Generative AI examples, (Panahi, Kivlin, and Louis 2024) illustrated a limited usage of the GPT model as a design review assistant tool that recommends design reviews based on previous projects and using computer vision and NLP techniques, and (Chou, Chong, and Liu 2024) developed a chatbot for flood risk management in dredging projects. In general, studies related to the application of Generative AI for NLP-based construction applications are rare, and this is a completely new research field for construction.

4 Application of NLP and Generative AI in Construction

The application of NLP and Generative AI in construction can not be summarised easily, but here, we briefly explored the most common areas in AEC where AI, mostly with the help of BIM, is utilized to automate different textual data analyses.

Risk management and safety: Case-based Reasoning (CBR) is an important AI approach in construction project risk management where previous knowledge and experience can be recalled and used as a starting point to solve new problems. Risk case databases often contain a huge amount of textual data and manual

reviewing, analyzing, and understanding. (H. Wang, Meng, and Zhu 2022) used BIM for lifecycle knowledge capture to address knowledge loss utilizing NLP and CBR. (Zou, Kiviniemi, and Jones 2017) reviewed the current literature on using NLP for the automatic retrieval and analysis of similar cases for construction project risk management. They also combined the Vector Space Model (VSM) and semantic query expansion methods to outline a framework for the Risk Case Retrieval System. The VSM could represent textual documents as vectors of identifiers, which could be used to compute the degree of similarity between documents and the query. Similarly, (Kim and Chi 2019) used the VSM and Tacit knowledge extraction model, which is a combination of rule-based and conditional random field (CRF) machine learning model for the extraction of tacit knowledge, to be used for one NLP-based accident case knowledge management system. A semantic search system was proposed by (Moon et al. 2018) to retrieve proper accident cases based on the user's deliberate intentions and to extract safety risk factors automatically by identifying the risk factors in construction accident reports. (T. Li and Harris 2019) investigated NLP for analyzing bridge inspection reports to find maintenance needed locations. (Ren and Zhang 2021) combined the power of NLP and Computer Vision to automatically monitor construction job site based on construction procedural documents. (Xu et al. 2021; Shen et al. 2022) used rule-based NLP techniques for safety risk identification in construction. (Yukang Wang et al. 2023) proposed to use NLP in BIM using industry foundation classes (IFC) for code compliance checking of fire safety aspects of geometry information benefiting from few tools available on the market for code compliance checking.

Contractual and legal assessment: (Khalef and El-adaway 2021) proposed a framework for combining NLP and Machine Learning classification for classifying and predicting substantial changes in contracts of Airport improvement program (AIP) projects. (Moon et al. 2021) benefiting from BiLSTM proposed a new NLP model for analysis of contracts for risk estimation. In another similar study, an automated system for reviewing construction specifications by analyzing the different semantic properties using NLP was proposed by (Moon, Lee, and Chi 2022). The authors made a semantic thesaurus from construction terms using Word2Vec embedding and then developed a NER model using LSTM. For automated schedule quality control in construction, a guidelines list called "logic" should be checked, and (Amer, Hockenmaier, and Golparvar-Fard 2022) proposed a deep learning model to learn scheduling domain knowledge from existing records completely. (Shooshtarian, Gurmu, and Sadick 2023) used different NLP techniques and models, including BERT, RoBERTa, fastText for analysis of building defects from legal documents. Staff assignment is part of the decision-making in construction projects, especially when responding to routine requests such as the change order and building service. However, the effectiveness is low due to manual processing by the management personnel. To improve the productivity of staff assignment, (Mo et al. 2020) proposed an NLP-based machine learning model that reads service request texts and automatically assigns workforce and priority.

Automatic compliance checking: The process of (ACC) usually has two inputs, one from regulatory documents and another from a regulatory-related queue from the user. (Zhang and El-Gohary 2016) has proposed a new method to extend the IFC schema that incorporates ACC-related information in an objective and semi-automated manner. They used NLP and ontology to capture the semantics of regulations and it contributes to alleviating the efforts to interpret regulations manually. (Zhang and El-Gohary 2016) developed a set of methods and combined them into one computational platform: (1) a method for concept extraction to extract regulatory concepts from regulatory documents, which utilizes POS patterns and pattern-matching-based rules; (2) a method for concept matching and semantic similarity assessment to select the most related IFC concepts to the extracted regulatory concepts, which utilizes term-based and semantic-based matching and a new equation for measuring semantic similarity; and (3) a method for relationship classification to predict the relationship between the extracted regulatory concepts and their most related IFC concepts.

Automatic underground utility compliance checking has been investigated by (S. Li, Cai, and Kamat 2016) using NLP and spatial reasoning. NLP algorithm translates textual descriptions of spatial configurations into spatial rules for geographical information systems (GIS). Spatial reasoning executes the extracted rules to identify noncompliance. Semi-automatic combinations of knowledge data from NLP in utility specification and spatial reasoning in GIS data provide a framework for utility compliance checking results (S. Li, Cai,

and Kamat 2016). (Song et al. 2018) proposed an NLP method for extraction of design requirements and semantic analysis of architectural properties and relations, from the textual architectural documents. (Al Qady and Kandil 2010) have used NLP to develop ontologies from construction contractual documents, NLP-based conceptual relationship recognition, and the shallow analysis method to automatically extract conceptual relationships from the text of contract documents. A Real-world Facility Installation Information Table (RFT) records information such as the building attribute names, installation locations, and number of facilities in a construction project (Xie et al. 2019). Using NLP, (Xie et al. 2019) proposed a matching method between RFT and BIM data.

Document analysis and classification: The process of automatically analyzing the content of documents and categorizing them according to the analysis results is called automated document classification (ADC). (Jung and Lee 2019) attempted to build a knowledge-based system for BIM projects called the Global BIM Dashboard. Based on their study, reviewing only one BIM-related document in detail takes approximately one to six hours. They utilized NLP and machine learning methods to perform automatic BIM-based ADC. A similar idea was proposed by (J. Wang et al. 2021) benefits from IFC specification. (Yang, Kim, and Kim 2021) used NLP-based Convolutional Neural Network (CNN) to automatically classify defect text data to overcome the limitations of managing defect cases through simple computerization. (Nabavi et al. 2023; Nabavi, Ramaji, and Sadeghi 2024) proposed a question-answering platform for extracting data from BIM using NLP. ifcOWL is a technology that provides a representation of the IFC schema using the Web Ontology Language (OWL). With the help of the ifcOWL ontology, it is now possible to represent building data using cutting-edge web technologies such as semantic web and linked data technologies. (Yin et al. 2024) proposed an ontology-based NLP approach to use in BIM utilizing IFC. They made an NLU-based OL method to enrich IFC ontology from BIM instance models automatically.

5 Conclusion and Future Path

As the gap analysis, (Chung et al. 2023) discussed that the first noteworthy observation in the analysis results of NLP tasks is the restricted number of data collection studies, specifically in the construction domain. The second discussion is that information extraction and retrieval mainly focused on entity-level information extraction, while the essential knowledge contained in the text data is often represented as the relationship between the entities. The third gap analysis of (Chung et al. 2023) is about the gap between computer science advancement in NLP and its application in construction. For instance, the recent ground-breaking introduction to ChatGPT is changing the way we formulate many NLP studies in construction as well as in computer science. Also, there are advances in question-and-answer (QA) tasks based on QA-specific model architectures (e.g., GPT and other generative models) and supportive resources (e.g., SQuAD, which is the standard Question Answering Dataset). Therefore, it is worth exploring the potential of language generation in construction. The gap analysis of (Chung et al. 2023) showed that the average gap between the NLP domain and the application of NLP in construction for traditional methods invested before 2010 is 22.5 years, while the average gap for methods between 2010 and 2021 is 4.4 years. One interesting discussion propounded by (Shamshiri, Ryu, and Park 2024) is that powerful BERT-based models are developed for domain-specific applications like health and law while the construction industry still lacks a fine-tuned BERT model tailored to construction-specific corpora.

Amongst all the domains of interest, safety management had the largest number of articles, with 42 publications, and Automated compliance checking was in the second rank, with 37 publications (Shamshiri, Ryu, and Park 2024). NLP is used for BIM-based cost analysis, quality control, facility management, and safety risk analysis. Research on the usage of NLP for intelligent query answering systems (QAS) and search engines to extract and retrieve information from BIM was considerable (Shamshiri, Ryu, and Park 2024). The review highlighted that reinforcement learning (RL) remains untapped in the construction field, while RL in NLP has made remarkable advancements. Another insight was that limited works are trying to use GPT models in BIM, while no scholars have yet attempted to use GPT4, Claude 2, and LLaMA families to explore their potential application in the construction sector. Figure 3 shows selective main topics from the literature review that NLP is used as a tool or there is a potential for NLP in future of construction 4.0.

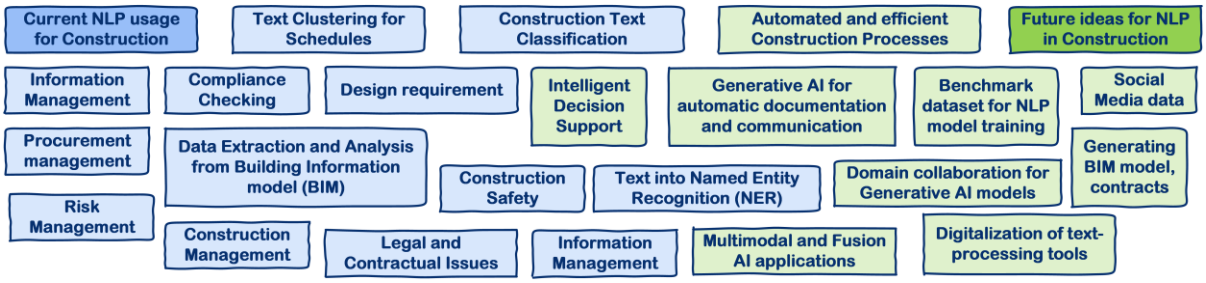

Figure 3. Selected topics and applications that NLP is used currently or could be used in construction, extracted from the literature.

6 References

Amer, Fouad, Julia Hockenmaier, and Mani Golparvar-Fard. 2022. "Learning and Critiquing Pairwise Activity Relationships for Schedule Quality Control via Deep Learning-Based Natural Language Processing." *Automation in Construction* 134. Elsevier B.V. doi:10.1016/j.autcon.2021.104036.

Chou, Jui Sheng, Pei Lun Chong, and Chi Yun Liu. 2024. "Deep Learning-Based Chatbot by Natural Language Processing for Supportive Risk Management in River Dredging Projects." *Engineering Applications of Artificial Intelligence* 131. Elsevier Ltd. doi:10.1016/j.engappai.2023.107744.

Chung, Sehwan, Seonghyeon Moon, Junghoon Kim, Jungyeon Kim, Seungmo Lim, and Seokho Chi. 2023. "Comparing Natural Language Processing (NLP) Applications in Construction and Computer Science Using Preferred Reporting Items for Systematic Reviews (PRISMA)." *Automation in Construction*. Elsevier B.V. doi:10.1016/j.autcon.2023.105020.

Ding, Yuexiong, Jie Ma, and Xiaowei Luo. 2022. "Applications of Natural Language Processing in Construction." *Automation in Construction*. Elsevier B.V. doi:10.1016/j.autcon.2022.104169.

Erfani, Abdolmajid, and Qingbin Cui. 2021. "Natural Language Processing Application in Construction Domain: An Integrative Review and Algorithms Comparison." In *Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021*, edited by R R A Issa, 26–33. American Society of Civil Engineers (ASCE). doi:10.1061/9780784483893.004.

Ghimire, Prashnna, Kyungki Kim, and Manoj Acharya. 2024. "Opportunities and Challenges of Generative AI in Construction Industry: Focusing on Adoption of Text-Based Models." *Buildings* 14 (1). Multidisciplinary Digital Publishing Institute (MDPI). doi:10.3390/buildings14010220.

Giuda, Giuseppe Martino Di, Mirko Locatelli, and Elena Seghezzi. 2020. "Natural Language Processing and Bim in Aeco Sector: A State of the Art." In *Proceedings of International Structural Engineering and Construction*, edited by H Askarinejad, S Yazdani, and A Singh, 7:CON-22-1-CON-22-6. ISEC Press. doi:10.14455/ISEC.2020.7(2).CON-22.

Hassan, Fahad Ul, and Tuyen Le. 2022. "State-of-the-Art Review on the Applicability of Natural Language Processing (NLP) Methods to Address Legal Issues in Construction." In *Construction Research Congress 2022: Project Management and Delivery, Controls, and Design and Materials - Selected Papers from Construction Research Congress 2022*, edited by F Jazizadeh, T Shealy, and M J Garvin, 3–C:159–68. American Society of Civil Engineers (ASCE). doi:10.1061/9780784483978.017.

Hassan, Fahad ul, Tuyen Le, and Xuan Lv. 2021. "Addressing Legal and Contractual Matters in Construction Using Natural Language Processing: A Critical Review." *Journal of Construction Engineering and Management* 147 (9). American Society of Civil Engineers (ASCE). doi:10.1061/(ASCE)CO.1943-7862.0002122.

Hong, Ying, Haiyan Xie, Gary Bhumbra, and Ioannis Brilakis. 2021. “Comparing Natural Language Processing Methods to Cluster Construction Schedules.” *Journal of Construction Engineering and Management* 147 (10). American Society of Civil Engineers (ASCE). doi:10.1061/(asce)co.1943-7862.0002165.

Jung, Namcheol, and Ghang Lee. 2019. “Automated Classification of Building Information Modeling (BIM) Case Studies by BIM Use Based on Natural Language Processing (NLP) and Unsupervised Learning.” *Advanced Engineering Informatics* 41 (August). Elsevier Ltd: 100917. doi:10.1016/j.aei.2019.04.007.

Khalef, Ramy, and Islam H. El-adaway. 2021. “Automated Identification of Substantial Changes in Construction Projects of Airport Improvement Program: Machine Learning and Natural Language Processing Comparative Analysis.” *Journal of Management in Engineering* 37 (6). American Society of Civil Engineers (ASCE). doi:10.1061/(asce)me.1943-5479.0000959.

Kilgarriff, Adam, and Christiane Fellbaum. 2000. “WordNet: An Electronic Lexical Database.” *Language* 76 (3). MIT press: 706. doi:10.2307/417141.

Kim, Taekhyung, and Seokho Chi. 2019. “Accident Case Retrieval and Analyses: Using Natural Language Processing in the Construction Industry.” *Journal of Construction Engineering and Management* 145 (3). American Society of Civil Engineers (ASCE). doi:10.1061/(ASCE)CO.1943-7862.0001625.

Li, Shuai, Hubo Cai, and Vineet R Kamat. 2016. “Integrating Natural Language Processing and Spatial Reasoning for Utility Compliance Checking.” *Journal of Construction Engineering and Management* 142 (12). American Society of Civil Engineers (ASCE). doi:10.1061/(asce)co.1943-7862.0001199.

Li, Tianshu, and Devin Harris. 2019. “Automated Construction of Bridge Condition Inventory Using Natural Language Processing and Historical Inspection Reports.” In *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIII 2019*, edited by A L Gyekenyesi, Yu T.-Y., H F Wu, and P J Shull, 10971:28. SPIE. doi:10.1117/12.2514006.

Liao, Wenjie, Xinzheng Lu, Yifan Fei, Yi Gu, and Yuli Huang. 2024. “Generative AI Design for Building Structures.” *Automation in Construction*. Elsevier B.V. doi:10.1016/j.autcon.2023.105187.

Locatelli, Mirko, Elena Seghezzi, Laura Pellegrini, Lavinia Chiara Tagliabue, and Giuseppe Martino Di Giuda. 2021. “Exploring Natural Language Processing in Construction and Integration with Building Information Modeling: A Scientometric Analysis.” *Buildings* 11 (12). MDPI. doi:10.3390/buildings11120583.

Mo, Yunjeong, Dong Zhao, Jing Du, Matt Syal, Azizan Aziz, and Heng Li. 2020. “Automated Staff Assignment for Building Maintenance Using Natural Language Processing.” *Automation in Construction* 113. Elsevier B.V. doi:10.1016/j.autcon.2020.103150.

Moon, Seonghyeon, Taekhyung Kim, Bon Gang Hwang, and Seokho Chi. 2018. “Analysis of Construction Accidents Based on Semantic Search and Natural Language Processing.” In *ISARC 2018 - 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things*. International Association for Automation and Robotics in Construction I.A.A.R.C). doi:10.22260/isarc2018/0109.

Moon, Seonghyeon, Gitaek Lee, and Seokho Chi. 2022. “Automated System for Construction Specification Review Using Natural Language Processing.” *Advanced Engineering Informatics* 51. Elsevier Ltd. doi:10.1016/j.aei.2021.101495.

Moon, Seonghyeon, Gitaek Lee, Seokho Chi, and Hyunchul Oh. 2021. “Automated Construction Specification Review with Named Entity Recognition Using Natural Language Processing.” *Journal of Construction Engineering and Management* 147 (1). American Society of Civil Engineers (ASCE). doi:10.1061/(asce)co.1943-7862.0001953.

Nabavi, Armin, Issa Ramaji, and Naimeh Sadeghi. 2024. “An Artificial Intelligence-Based Framework for Automated Information Inquiry from Building Information Models Using Natural Language Processing

and Ontology." In *Computing in Civil Engineering 2023: Visualization, Information Modeling, and Simulation - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023*, edited by Y Turkan, J Louis, F Leite, and S Ergan, 381–89. American Society of Civil Engineers (ASCE). doi:10.1061/9780784485231.046.

Nabavi, Armin, Issa Ramaji, Naimeh Sadeghi, and Anne Anderson. 2023. "LEVERAGING NATURAL LANGUAGE PROCESSING FOR AUTOMATED INFORMATION INQUIRY FROM BUILDING INFORMATION MODELS." *Journal of Information Technology in Construction* 28. International Council for Research and Innovation in Building and Construction: 266–85. doi:10.36680/J.ITCON.2023.013.

Panahi, Roshan, John Paul Kivlin, and Joseph Louis. 2024. "Request for Information (RFI) Recommender System for Pre-Construction Design Review Application Using Natural Language Processing, Chat-GPT, and Computer Vision." In *Computing in Civil Engineering 2023: Data, Sensing, and Analytics - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023*, edited by Y Turkan, J Louis, F Leite, and S Ergan, 159–66. American Society of Civil Engineers (ASCE). doi:10.1061/9780784485224.020.

Peng, Xiangshu, Zhiming Ma, Ping Wang, Yaoxian Huang, and Lixia Zhang. 2023. "Research on Natural Language Processing Technology in Construction Project Management." In *4th International Conference on Artificial Intelligence in China, AIC 2022*, edited by Liang Q., Wang W., Mu J., Liu X., and Na Z., 871 LNEE:11–17. School of Computer Science and Technology, Xinjiang Normal University, Urumchi, 830054, China: Springer Science and Business Media Deutschland GmbH. doi:10.1007/978-981-99-1256-8_2.

Pham, Hieu T. T. L., and SangUk Han. 2023. "Natural Language Processing with Multitask Classification for Semantic Prediction of Risk-Handling Actions in Construction Contracts." *Journal of Computing in Civil Engineering* 37 (6). American Society of Civil Engineers (ASCE). doi:10.1061/jccee5.cpeng-5218.

Qady, Mohammed Al, and Amr Kandil. 2010. "Concept Relation Extraction from Construction Documents Using Natural Language Processing." *Journal of Construction Engineering and Management* 136 (3): 294–302. doi:10.1061/(ASCE)CO.1943-7862.0000131.

Ren, Ran, and Jiansong Zhang. 2021. "An Integrated Framework to Support Construction Monitoring Automation Using Natural Language Processing and Sensing Technologies." In *Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021*, edited by R R A Issa, 1101–9. American Society of Civil Engineers (ASCE). doi:10.1061/9780784483893.135.

Shamshiri, Alireza, Kyeong Rok Ryu, and June Young Park. 2024. "Text Mining and Natural Language Processing in Construction." *Automation in Construction*. Elsevier B.V. doi:10.1016/j.autcon.2023.105200.

Shen, Qiyu, Songfei Wu, Yichuan Deng, Hui Deng, and Jack C.P. Cheng. 2022. "BIM-Based Dynamic Construction Safety Rule Checking Using Ontology and Natural Language Processing." *Buildings* 12 (5). MDPI. doi:10.3390/buildings12050564.

Shooshtarian, Salman, Argaw Tarekegn Gurmu, and Abdul Manan Sadick. 2023. "Application of Natural Language Processing in Residential Building Defects Analysis: Australian Stakeholders' Perceptions, Causes and Types." *Engineering Applications of Artificial Intelligence* 126. Elsevier Ltd. doi:10.1016/j.engappai.2023.107178.

Song, Jae Yeol, Jin Sung Kim, Hayan Kim, Jungsik Choi, and Jin Kook Lee. 2018. "Approach to Capturing Design Requirements from the Existing Architectural Documents Using Natural Language Processing Technique." In *CAADRIA 2018 - 23rd International Conference on Computer-Aided Architectural Design Research in Asia: Learning, Prototyping and Adapting*, edited by S Alhadidi, T Fukuda, W

Huang, P Janssen, and K Crolla, 2:247–54. The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA). doi:10.52842/conf.caadria.2018.2.247.

Wang, Hao, Xianhai Meng, and Xingyu Zhu. 2022. “Improving Knowledge Capture and Retrieval in the BIM Environment: Combining Case-Based Reasoning and Natural Language Processing.” *Automation in Construction* 139. Elsevier B.V. doi:10.1016/j.autcon.2022.104317.

Wang, Jia, Xinao Gao, Xiaoping Zhou, and Qingsheng Xie. 2021. “Multi-Scale Information Retrieval for BIM Using Hierarchical Structure Modelling and Natural Language Processing.” *Journal of Information Technology in Construction* 26. International Council for Research and Innovation in Building and Construction: 409–26. doi:10.36680/j.itcon.2021.022.

Wang, Yu, Hui Deng, Xiao Yao Li, and Yi Chuan Deng. 2020. “A review of natural language processing application in construction engineering.” *Journal of Graphics* 41 (4). Editorial of Board of Journal of Graphics: 501–11. doi:10.11996/JG.j.2095-302X.2020040501.

Wang, Yukang, Yue Liu, Haozhe Cai, Jia Wang, and Xiaoping Zhou. 2023. “An Automated Fire Code Compliance Checking Jointly Using Building Information Models and Natural Language Processing.” *Fire* 6 (9). Multidisciplinary Digital Publishing Institute (MDPI). doi:10.3390/fire6090358.

Wu, Chengke, Xiao Li, Yuanjun Guo, Jun Wang, Zengle Ren, Meng Wang, and Zhile Yang. 2022. “Natural Language Processing for Smart Construction: Current Status and Future Directions.” *Automation in Construction*. Elsevier B.V. doi:10.1016/j.autcon.2021.104059.

Xie, Qingsheng, Xiaoping Zhou, Jia Wang, Xinao Gao, Xi Chen, and Liu Chun. 2019. “Matching Real-World Facilities to Building Information Modeling Data Using Natural Language Processing.” *IEEE Access* 7. Institute of Electrical and Electronics Engineers Inc.: 119465–75. doi:10.1109/ACCESS.2019.2937219.

Xu, Na, Ling Ma, Li Wang, Yongliang Deng, and Guodong Ni. 2021. “Extracting Domain Knowledge Elements of Construction Safety Management: Rule-Based Approach Using Chinese Natural Language Processing.” *Journal of Management in Engineering* 37 (2). School of Mechanics and Civil Engineering, China Univ. of Mining and Technology, Xuzhou, 221000, China: American Society of Civil Engineers (ASCE): 04021001. doi:10.1061/(ASCE)ME.1943-5479.0000870.

Yang, Donguk, Byeol Kim, and Hayoung Kim. 2021. “Automated Defect Classification in the Maintenance Phase Using a Channel Attention-Based Convolutional Neural Network Model of Natural Language Processing.” *International Journal of Sustainable Building Technology and Urban Development* 12 (2). Sustainable Building Research Center: 96–109. doi:10.22712/susb.20210009.

Yin, Mengtian, Llewellyn Tang, Chris Webster, Xiaoyue Yi, Huaquan Ying, and Ya Wen. 2024. “A Deep Natural Language Processing-based Method for Ontology Learning of Project-specific Properties from Building Information Models.” *Computer-Aided Civil and Infrastructure Engineering* 39 (1). John Wiley and Sons Inc: 20–45. doi:10.1111/mice.13013.

Zhang, Jiansong, and Nora M El-Gohary. 2016. “Extending Building Information Models Semiautomatically Using Semantic Natural Language Processing Techniques.” *Journal of Computing in Civil Engineering* 30 (5). American Society of Civil Engineers (ASCE). doi:10.1061/(asce)cp.1943-5487.0000536.

Zheng, Zhe, Xin Zheng Lu, Ke Yin Chen, Yu Cheng Zhou, and Jia Rui Lin. 2022. “Pretrained Domain-Specific Language Model for Natural Language Processing Tasks in the AEC Domain.” *Computers in Industry* 142. Elsevier B.V. doi:10.1016/j.compind.2022.103733.

Zou, Yang, Arto Kiviniemi, and Stephen W Jones. 2017. “Retrieving Similar Cases for Construction Project Risk Management Using Natural Language Processing Techniques.” *Automation in Construction* 80. Elsevier B.V.: 66–76. doi:10.1016/j.autcon.2017.04.003.