Bridging Building Information Modelling and Natural Language Processing: A Systematic Review on Current Applications and Limitations

Tessa Marie Oberhoff[1], Julian Cloos[1], Sven Mackenbach[1] and Katharina Klemt-Albert[1]

[1]Institute for Construction Management, Digital Engineering and Robotics in Construction
RWTH Aachen University, Germany
oberhoff@icom.rwth-achen.de; julian.cloos1@rwth-aachen.de; mackenbach@icom.rwth-aachen.de; klemt-albert@icom.rwth-aachen.de

Abstract –

Building Information Modelling (BIM) is widely recognised as a key driver of digital transformation in the architecture, engineering and construction industry. However, its full potential remains underutilised, partly due to the vast and complex nature of BIM-related data. At the same time, Natural Language Processing (NLP), a rapidly evolving field of artificial intelligence, offers advanced methods for interpreting and leveraging text-based information. The convergence of these two technologies holds considerable potential to facilitate human interaction with BIM data by making it more accessible. Despite this potential, research directly linking NLP and BIM remains limited. This paper attempts to bridge this gap by mapping and evaluating the current state of research on NLP applications within BIM workflows. To this end, a Systematic Literature Review is conducted in accordance with the methodological framework proposed by Fink, identifying 92 relevant publications. The metadata of these publications was analysed using VOSviewer, followed by a detailed content analysis. For each source, the BIM-related tasks optimised by NLP and NLP functionality applied to BIM data are identified, and a novel approach is proposed to classify the BIM-NLP integration into five categories: Extract, Query, Enrich, Modify, and Match. The analysis reveals that most existing literature explores NLP application in a peripheral or exploratory manner, with few instances of deep integration into BIM workflows. Nevertheless, the reviewed literature highlights the potential of NLP to improve data accessibility, automate information processing, and enhance human-machine interaction.

Keywords –

BIM–NLP integration, Large Language Models, Human–Machine Interaction, Data Accessibility

1 Introduction

The architecture, engineering, and construction (AEC) industry is a highly data-intensive domain. Throughout the entire building lifecycle, large volumes of heterogeneous information are generated and exchanged. Despite ongoing digitisation and the increasing adoption of Building Information Modelling (BIM), a substantial share of information exchange in the AEC sector continues to rely on textual information, from both unstructured documents and semi-structured but fragmented information systems. [1–3]

Natural Language Processing (NLP) provides methods for the automated analysis of such textual information [4, 5]. Recent advances in NLP, particularly the emergence of Large Language Models (LLMs), have significantly expanded the ability to capture semantic meaning, contextual relationships, and domain-specific knowledge from vast and unstructured text sources [6, 7]. In the context of the transition toward a data-driven and model-based construction industry, the combination of BIM and NLP has been identified as a technology with considerable potential [8–10]. BIM relies on structured information models that are commonly exchanged using the Industry Foundation Classes (IFC) standard, which is based on formal, machine-readable text [11]. Integrating NLP techniques with BIM, therefore, enables the linkage of textual information from other sources within one central information model of a built asset. This provides the potential for more data-driven processes, which can result in enhanced decision-making and greater automation across the entire lifecycle of a building.

To understand the potential of BIM–NLP integration current applications and limitations need to be analysed. But only a limited number of studies have explicitly reviewed the combined application of BIM and NLP. Di Giuda et al. [12] presented a literature review, identifying several potential application areas of NLP in AEC, but with a strong focus on using NLP for requirements engineering at an early project phase to initialize BIM. Two subsequent studies by Locatelli et al. [2, 13] applied scientometric methods to analyse BIM–NLP integration. These analyses identified a rapidly increasing scientific interest in NLP and Semantic BIM as independent research domains and clustered the identified articles thematically. However, the clusters lack a unified clustering strategy, consisting of a mix of BIM-related tasks, NLP tasks, and review articles.

Overall, the existing review studies indicate growing academic interest in the integration of BIM and NLP but also reveal unresolved research gaps. In particular, there is a lack of a systematic, content-oriented literature review that critically examines NLP integration into BIM-centred workflows. In addition, due to rapid developments in Artificial Intelligence (AI) methods, NLP is a dynamically evolving field [5]. Consequently, the findings of existing studies are no longer up to date. Hence, the objective of this article is to conduct a Systematic Literature Review on the combined use of BIM and NLP in the AEC sector. The study aims to approach existing research from a content perspective, identify dominant application areas, and assess current limitations to outline directions for future research.

2 Theoretical Background

2.1 Natural Language Processing

NLP is a technique to analyse, understand, and work with human language with the help of computers [14]. NLP involves converting unstructured language data into a structured data format. This enables machines to identify and understand language and text to generate relevant responses [15]. Contrary to some definitions, NLP does not constitute a subfield of AI per se; however, it does utilise AI methods [4, 5], as shown in a schematic classification of different forms of NLP in Figure 1.

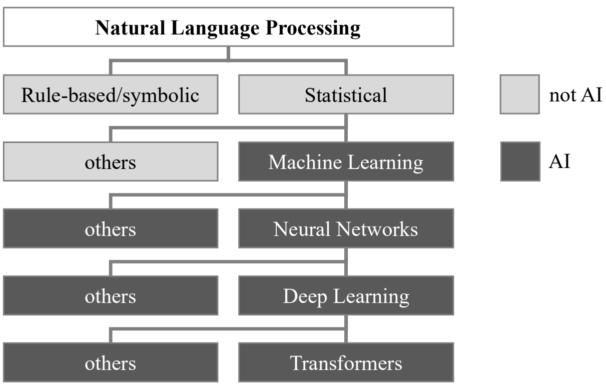

Figure 1. Schematic classification of NLP methods and models based on Canchila et al. [5]

According to Locatelli et al. [2] the first scientific article about NLP application in the field of AEC was published in 1989 by Logcher et al., in the form of an article bearing the title "Knowledge Processing for Construction Management Data Base"[16]. Since then, methods have evolved from rule-based/symbolic procedures, such as ontologies, to statistical procedures. In the field of statistical methods, it is possible to distinguish between classical statistical methods and machine learning methods, which are part of the field of AI. Within machine learning, one can differentiate between Neural Networks (NN) and other methods. Continuing along the chain, among NN methods, Deep Learning is the subfield leading to Transformer architectures, which have attracted considerable attention for several years [15]. Transformer models refer to a NN architecture introduced by Vaswani et al. [17]. It is a general-purpose architecture that can be applied to many tasks beyond language. Most contemporary LLMs (e.g. GPT-5, LLaMa, and BERT) are implemented using Transformer architectures [5]. LLMs, in turn, are models defined by their large scale in terms of parameters, training data, and computational resources, as well as their ability to perform a wide range of language-related tasks [18].

2.2 Building Information Modelling

BIM is a model-based methodology for the digital representation and management of physical and functional characteristics of built assets [19]. BIM enables the integration of geometric, semantic, and relational information into a coherent information model that can be used across different phases of the building lifecycle [19]. In this context, the 3D geometric model is commonly extended by additional information, so-called dimensions. Typically, temporal information to support construction sequencing and scheduling is referred to as 4D-BIM, and adding cost-related data to enable quantity-based cost estimation and financial control is referred to as 5D-BIM [20]. Beyond these dimensions, BIM models can be further enriched with information related to aspects such as sustainability, safety, or facility management [20]. The extension of BIM into multiple dimensions enables a more comprehensive and integrated management of construction projects by combining heterogeneous information within the digital model of the built asset. A central aspect of storing the relevant information is the underlying data structure of the BIM model. BIM models are typically exchanged using the IFC standard, which provides a formal, object-oriented schema for representing building components, their properties, and their relationships [11, 21]. IFC is based on machine-readable and hierarchical data structures that support interoperability between software systems [22]. These structured representations allow information to be queried, validated, and processed algorithmically, forming the basis for NLP applications [11].

3 Methodology

To identify relevant publications, a Systematic Literature Review according to the methodology of Fink [23] is conducted. The first step in this methodology is defining research questions (RQs). In this article, three RQs are posed. The first RQ pertains to the identification of relevant literature and the initial characterisation of the same. The specific inquiry is as follows:

- RQ1: What scientific articles exist on the integration of BIM and NLP?

Following the next steps according to Fink [23], subsequently the databases for the search are defined. The databases Engineering Village (Compendex), IEEE Xplore, Scopus, Taylor & Francis, Wiley Online Library, and Web of Science are used, since they cover a large part of common databases for both general and engineering-specific topics. Additionally, the DBs allow suitable options for keyword entry, as well as filter and export functions. The search string for the Systematic Literature Review consists of one component on the topic of "BIM" and a second component on the topic of "NLP". Synonyms for the two main search words are additionally collected. In the NLP component, "LLMs" are included as the only specific method, as they have gained particular prominence in recent years. The synonyms are linked by the Boolean operator "OR" and the two components are linked by an "AND" operator.

This procedure identifies 430 articles, from which duplicates are first removed. According to Fink [23], the remaining articles are then subjected to a first and second filtering procedure (screening). Formal quality criteria are applied in the first filtering process. These quality criteria include:

- Article not accessible
- Article no longer available
- Article not in German or English

Material quality criteria are applied in the second filtering process. These include, above all, thematic relevance to the research question. For instance, reasons for exclusion include:

- Article does not relate to the construction industry
- Article does not relate to the use of NLP
- Article only mentions BIM in the outlook/introduction

The review process is shown in Figure 2, based on a PRISMA diagram [24]. It shows the number of articles included or excluded in each step, along with the reason for exclusion.

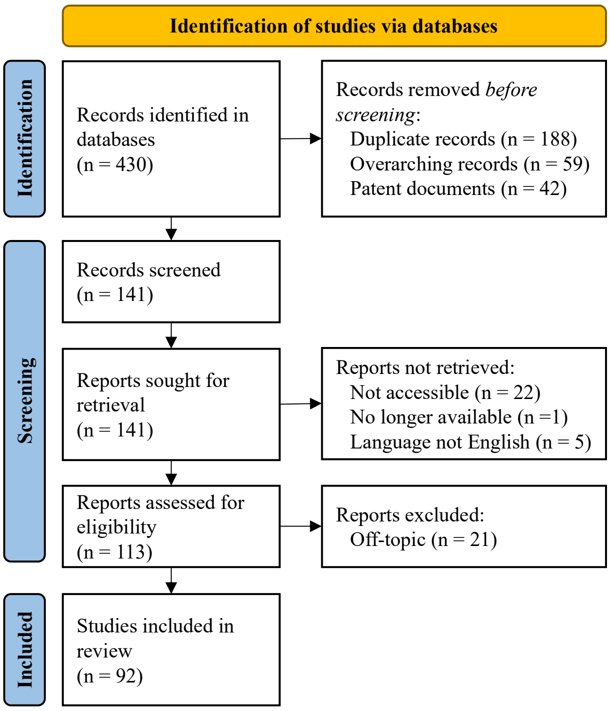

Figure 2. Conducted Systematic Literature Review process based on a PRISMA-diagram

As a result of the Systematic Literature Review, 92 articles remain for further analysis. To answer the first RQ, scientometric data is evaluated by VOSviewer [25]. This paper highlights two aspects: the year of publication and a classification of the publication type. The development of publication frequency indicates the interest in a research field and can show trends. Analysing the publication type, differentiating between "Original Research", "Reviews", and "Case Studies", can give a first impression of the content.

The bibliometric data is followed by an in-depth content analysis of the BIM–NLP applications described in the articles. Therefore, two more RQs are defined:

- RQ2: Which BIM workflow-related tasks are tackled by NLP integration in the identified articles?
- RQ3: Which NLP functionalities are used regarding BIM data in the identified articles?

To answer both RQs, clustering techniques are applied. For RQ2, the clusters are formed based on the BIM-related tasks within a building's lifecycle. At this point, the primary focus is on the identification of tasks within the BIM workflow that are impacted, while the specific functionality of NLP is not a primary concern. The task clusters are developed based on existing review research articles, and are then adapted to or supplemented by novel categories. For RQ3, the clusters are formed based on the NLP functionality applied to BIM data. These clusters are not related to specific NLP methods or models, but rather characterise the actions performed by the NLP application on the BIM data. Therefore, new categories are developed.

4 Analysis

4.1 Year of Publication

Figure 3 shows how the identified publications are distributed over time. The diagram starts in 2014, as no articles published before this date were identified.

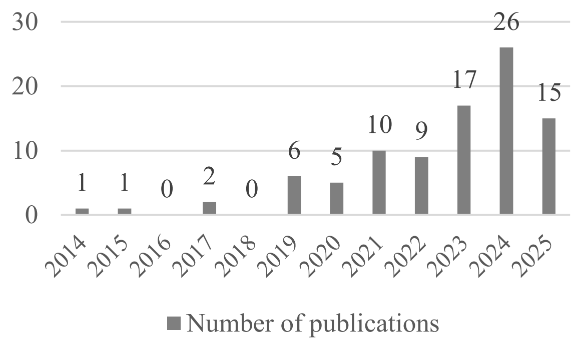

Figure 3. Development of the number of publications about BIM–NLP integration

It should be noted that the number of articles for 2025 only covers part of the year, as the research was conducted in May 2025. Based on the results from May, a higher number of publications can be expected for 2025 than for 2024. The development of publication numbers shows a rising research interest in the integration of BIM and NLP, confirming the findings of the reviews mentioned in Chapter 1. At the same time, the number of existing studies is still rather limited.

4.2 Type of Publication

The following section provides a concise overview of the categorisation of each publication into one or more types of publication. The results of the publication type analysis are shown in Table 1.

Table 1. Number and percentage of articles per publication type

Publication type	Number	% (N=92)
Original Research	78	84
Review	11	12
Case Study	3	3

The majority of the identified publications are classified as Original Research. These contributions investigate a wide range of applications related to the integration of BIM and NLP, typically focusing on specific methods, tools, or workflows. Given the diversity of application contexts and technical approaches, a detailed content-oriented analysis of these studies is provided in Chapters 4.3 and 4.4.

The second largest group consists of Reviews. Three of these publications [2, 12, 13], briefly introduced in Chapter 1, focuses explicitly on the combined application of BIM and NLP. One additional review addresses the use of NLP in the context of construction schedule management [9]. The remaining Reviews cover related but more specialized topics, including reviews of computational techniques supporting specific BIM-related tasks, such as quantity take-off, knowledge management, and compliance checking, as well as broader technological perspectives on digital twins, big data analytics, and AI in the AEC industry. Two reviews focus on specific NLP-related models or techniques, namely GPT models (short for: Generative Pre-trained Transformer models; a special kind of Transformer model) and Deep Learning approaches, while one addresses digital twin applications in tunnel construction.

Only a small number of publications are classified as pure Case Study. This limited representation indicates that, to date, relatively few BIM–NLP applications have been empirically evaluated in real-world settings. The field, therefore, currently remains largely exploratory, and there is an evident need for further validation of proposed approaches through practical implementations.

4.3 BIM Task Classification

This section provides a concise overview of the BIM workflow-related task clusters identified in the relevant publications. The results of the clustering are demonstrated in Table 2. The total number of clusters is nine. Seven publications, all Reviews, were excluded from this analysis as they deal with various BIM task clusters.

Table 2. Number and percentage of articles per BIM task cluster

BIM task cluster	Number	% (N=85)
Knowledge Management	23	27
Compliance Checking	16	19
BIM Authoring	12	14
Design and Planning	10	12
Facility Management	6	7
Risk and Safety Management	6	7
Cost Management	3	4
Robot Control	3	4
Schedule Management	3	4
Sustainability Assessment	3	4
various	7	-

Knowledge Management: This category comprises the capture, structuring, and reuse of project-related knowledge, such as information from previous projects, guidelines, and experiential knowledge. Its purpose is to systematically provide knowledge to support and enhance other BIM-related tasks. The cluster is only assigned if the activities are not related to any of the following specific BIM task clusters. In previous review articles, the cluster was named differently, e.g. Information management.

Compliance Checking: This category refers to the verification of BIM models against regulatory requirements, standards, codes, and project-specific rules. It encompasses both formal rule-based checks and content-related compliance assessments. The cluster has the second highest number of assigned publications and was already mentioned in all previous reviews on BIM–NLP integration. This indicates that Compliance checking is one of the most well-established BIM workflow-related tasks, where NLP application is studied.

BIM Authoring: This category covers the creation, modification, and continuous maintenance of BIM models throughout the project lifecycle. It includes model structuring, attribute management, and information enrichment. In previous studies, only information enrichment of BIM models was mentioned. This shows that BIM modelling from scratch is new, but, as the high number of publications indicates, it is a topic of rising interest.

Design and Planning: This category focuses on supporting decisions during early project phases. It includes design alternative exploration, scenario analysis, and generative approaches that support decision-making. A similar task cluster, namely Design management, was identified in one previous study. The name of the cluster was adapted to focus on the design process and illustrate the difference from the cluster Compliance Checking.

Facility Management: This category describes the use of BIM information to support building operation after project completion. It includes activities related to maintenance, operation, and asset management, mostly related to decision support during those stages. The task cluster was only mentioned by the more recent of the previous studies and seems to be a rather slowly developing application of NLP.

Risk and Safety Management: This category addresses the identification, assessment, and prevention of risks and safety issues in a BIM context. Most publications in this cluster focus on the construction phase. Previous studies also mentioned risk and safety-related applications. Partly, publications related to this BIM task cluster were excluded during the Systematic Literature Review, due to the NLP application often focusing on safety reports as separate documents, as opposed to being incorporated into BIM workflows.

Cost Management: This category addresses activities related to cost estimation, cost planning, and cost control based on 5D BIM. As the number of publications shows, it is currently a rather small research field.

Robot Control: This category covers the application of robotic and automated systems in construction and building operation contexts. It includes automated execution activities and interactions between humans and robots in a BIM context.

Schedule Management: This category involves the planning, monitoring, and optimization of 4D BIM. It includes construction sequencing, progress tracking, and coordination of resources. Even though the number of assigned publications is relatively small, one of the Reviews is related to the BIM task cluster Schedule Management.

Sustainability Assessment: This category focuses on the evaluation of buildings across their lifecycle, especially in the context of Life Cycle Assessment (LCA). It intentionally does not include design recommendations to improve the sustainability of the building, nor does it consider energy optimization decision-making, as both activities are included in other BIM task clusters.

4.4 NLP Functionality Classification

In order to explore the functions performed by NLP in more detail, this section examines the functionality of NLP methods with BIM data, independent of the concrete BIM task supported. The NLP functionality clusters intentionally avoid using a hierarchical structure. To characterize the functionality, five distinct clusters were developed. The resulting clusters and the number and percentage of assigned publications are demonstrated in Table 3 and discussed following the table. All eleven publications classified as Review were excluded from the analysis, as they deal with various NLP functionality clusters. The remaining publications can be assigned to more than one NLP cluster, as some of the presented approaches incorporate more than one NLP application.

Table 3. Number and percentage of articles per NLP functionality cluster

NLP functionality cluster	Number	% (N=81)
Extract	32	40
Query	31	38
Match	20	25
Enrich	17	21
Modify	12	15
various	11	-

Extract: This category refers to approaches in which NLP is applied to derive structured or semi-structured information from BIM-related sources, such as regulatory texts, technical specifications, or reports. The extracted information is typically transformed into formats suitable for further processing and use in BIM-based workflows. The BIM database itself is not analysed or changed by NLP methods in this cluster. Extract has the highest number of allocated publications. This may well be the case, given that NLP was originally developed to work with human language rather than machine-readable texts, such as IFC models.

Query: This category includes methods that allow users to access BIM-related information through natural language queries. NLP is employed to interpret user inputs and retrieve relevant data from BIM models, databases, or linked information systems, often via conversational or assistant-based interfaces, without altering the underlying model. Chatbots are identified to be a trending research topic in this cluster, leading to the second highest number of allocated publications.

Match: This category comprises methods that establish correspondences between BIM data and external information sources, such as regulations, schedules, databases, or real-world entities. NLP supports the alignment and comparison of heterogeneous representations, enabling activities such as data integration and cross-system mapping. Thus, the IFC file of the BIM model is unaffected.

Enrich: This category encompasses techniques that use NLP to augment BIM data with additional semantic information. Such approaches add or infer attributes, classifications, relationships, or ontological links, thereby improving the semantic completeness and interpretability of BIM models and associated datasets. BIM data is enhanced by adding information, but no previously existing data is changed or deleted.

Modify: This category covers approaches in which NLP is used to directly influence or perform changes to BIM data. It includes text- or speech-driven creation and modification of model elements, execution of modelling commands, and automated or semi-automated generation of BIM content. The limited number of publications assigned to this cluster could indicate the novelty of these NLP functionalities or the complexity, resulting from the profound impact on the BIM data structure in this cluster.

5 Discussion

The analysis reveals distinct patterns regarding both the BIM-related tasks addressed and the functionalities for which NLP is integrated into BIM-based workflows.

Among the BIM task clusters, Knowledge Management represents the largest share of publications. This finding reflects the strong reliance of BIM processes on textual information, such as requirements, reports, guidelines, and project documentation, which can be processed effectively using NLP techniques. The predominance of this category is consistent with earlier observations [12, 13]. Compliance Checking constitutes the second most frequent task cluster. The review shows that compliance-related studies often combine information extraction from regulatory texts with alignment or comparison against BIM data, indicating a close relationship between the Compliance Checking task cluster and the NLP functionality clusters Extract and Match. Tasks related to the clusters BIM Authoring and Design and Planning are also well represented. A shift toward more interactive and model-centred applications can be observed. Similar tendencies are discussed in one of the identified review articles [7]. In contrast, task clusters that typically depend on numerical/geometric data or require tight integration with domain-specific tools remain limited. This suggests that current methods are less applicable to such data-intensive tasks and are less applicable in complex BIM ecosystems.

Regarding NLP functionality clusters, the categories Extract and Query are most prevalent, followed by Match. The strong presence of Extract and Query is consistent with the dominance of the BIM task clusters Knowledge Management and Compliance Checking, both of which primarily involve handling textual information. The comparatively lower frequency of Enrich and Modify shows that semantic augmentation and model-altering applications remain limited. Extract and Query functionalities may dominate, due to requiring only limited interaction with BIM schemas, whereas the more advanced functionalities Modify and Enrich remain comparatively rare due to their need for strict compliance with structured data models and semantic consistency. This discrepancy can be interpreted as a representation mismatch problem: NLP systems operate on probabilistic, context-dependent representations, while BIM relies on deterministic, schema-constrained data structures. Bridging this mismatch is therefore a broader integration problem between statistical and symbolic systems.

To enable a transition from passive NLP interaction, such as querying, to active model modification, this study proposes several technical prerequisites. First, the development of schema-aware language models is essential, as current LLMs lack an explicit understanding of BIM (especially IFC) schemas. Future research should therefore focus on fine-tuning models on IFC-based datasets. Second, formal validation layers are required to ensure the correctness of model modifications. This includes the integration of rule-based validation systems for IFC compliance, as well as hybrid pipelines that combine NLP outputs with symbolic reasoning. Without such mechanisms, NLP-generated modifications risk producing invalid or inconsistent BIM models. [26]

The analysis also indicates significant barriers to real-world adoption, as reflected in the low number of case studies identified in the literature. One possible reason is data privacy and security concerns, as construction projects often involve sensitive contractual and design information [27]. Additionally, issues of trust play a crucial role, as practitioners require transparent and verifiable outputs due to the high impact of modifications on BIM data, which the black-box nature of many AI-based NLP techniques precludes. Inaccuracies can propagate into design decisions, cost estimation, or construction execution, leading to considerable financial risks [28]. Instead of fully automated pipelines, human-machine-interaction, like Explainable AI and human-in-the-loop approaches are therefore seen as promising [29]. This trend can also be observed in the review, namely in the rising number of co-pilot systems.

Overall, the results indicate that current research mainly applies NLP at the interface or preprocessing level, while deeper integration into BIM workflows remains uncommon. Although recent publications demonstrate technically advanced prototypes, particularly in connection with Transformer architectures, their scope is often limited to controlled scenarios and lacks validation in complex, real-world projects.

6 Conclusion

This study presents a Systematic Literature Review on the integration of BIM and NLP, identifying 92 relevant publications and analysing them with respect to BIM-related tasks and NLP functionalities. The results confirm that BIM–NLP integration is an emerging but still fragmented research field. To characterise the nature of BIM–NLP integration, this study introduces a novel classification of NLP functionalities in connection with BIM data, consisting of the five clusters Extract, Query, Enrich, Modify, and Match. The analysis shows that most existing work focuses on information extraction and querying, whereas deeper integration forms, such as semantic enrichment and model modification, are less common. The discussion identifies barriers that hinder deeper integration of NLP into BIM workflows. A key finding is the mismatch between probabilistic NLP models and schema-based BIM representations, which limits the ability of current approaches to directly modify or enrich BIM data. Furthermore, data privacy concerns in handling sensitive project information, and a lack of trust due to limited reliability of NLP outputs obstruct real-world adoption.

Some limitations of this review should be acknowledged. The literature search was conducted up to May 2025, which restricts the coverage of very recent developments in a rapidly evolving field. In addition, the clustering of BIM-related tasks and NLP functionality involves interpretative decisions, particularly for studies addressing multiple objectives.

To bridge the gap between NLP and BIM, future research should focus on the development of schema-aware language models and NLP–rule-based validation frameworks including technical and human quality assurance. In addition, the evaluation of approaches in real-world scenarios is a high-priority step toward practical adoption. Ultimately, the integration of NLP into BIM has the potential to improve automation of complex information processes and enhance decision-making in a BIM context. However, realising this potential requires solutions that can operate reliably within existing BIM workflows.

References

[1] F. Opitz, R. Windisch, and R. J. Scherer. Integration of Document- and Model-based Building Information for Project Management Support. *Procedia Engineering*, 85:403–411, 2014. doi: 10.1016/j.proeng.2014.10.566.

[2] M. Locatelli, E. Seghezzi, L. Pellegrini, L. C. Tagliabue, and G. M. Di Giuda. Exploring natural language processing in construction and integration with building information modeling: A scientometric analysis. *Buildings*, 11(12):583, 2021. doi: 10.3390/buildings11120583.

[3] A. Shamshiri, K. R. Ryu, and J. Y. Park. Text mining and natural language processing in construction. *Automation in Construction*, 158:105200, 2024. doi: 10.1016/j.autcon.2023.105200.

[4] C. Feng, Y. Li, Z. Chen, and L. Guo. The Evolution and Breakthrough of Natural Language Processing: The Revolution from Rules to Deep Learning. In *Proceedings of the 5th International Conference on Computer Science and Management Technology*, Xiamen Guangdong, China, 2024. doi: 10.1145/3708036.3708089.

[5] S. Canchila, C. Meneses-Eraso, J. Casanoves-Boix, P. Cortés-Pellicer, and F. Castelló-Sirvent. Natural language processing: An overview of models, transformers and applied practices. *ComSIS*, 21(3): 1097–1145, 2024. doi: 10.2298/CSIS230217031C.

[6] X. Xiahou, G. Chen, Z. Li, X. Xu, and Q. Li. Knowledge Management in Construction Quality Management: Current State, Challenges, and Future Directions. *IEEE Transactions on Engineering Management*. 72:1069–1088, 2025. doi: 10.1109/TEM.2025.3550354.

[7] A. Saka, R. Taiwo, N. Saka, B. Salami, S. Ajayi, K. Akande, and H. Kazemi. GPT models in construction industry: Opportunities, limitations, and a use case validation. *Developments in the Built Environment*, 17:100300, 2024. doi: 10.1016/j.dibe.2023.100300.

[8] E. Ohene, G. Nani, M. F. Antwi-Afari, A. Darko, L. A. Addai. and E. Horvey. Big data analytics in the AEC industry: scientometric review and synthesis of research activities. *Engineering Construction &*

Architectural Management. 32(11):7299–7331, 2025. doi: 10.1108/ECAM-01-2024-0144.
[9] A. K. Singh, A. Pal, P. Kumar, J. J. Lin, and S.-H. Hsieh. Prospects of Integrating BIM and NLP for Automatic Construction Schedule Management. In *41st International Symposium on Automation and Robotics in Construction*, pages 238–245, Lille, France, 2024. doi: 10.22260/ISARC2023/0034.
[10] A. Corneli, L. Binni, F. Spegni, L. Messi, and B. Naticchia, NATURAL LANGUAGE PROCESSING FOR CONSTRUCTION SITES MANAGEMENT. In *Proceedings of the 2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference*, Heraklion, Crete, Greece, 2023. doi: 10.35490/EC3.2023.272.
[11] M. Selim, K. Nassar, and O. Hosny, Hg., Automating BIM (IFC) Data Analysis Using LangGraph. In *Proceedings of the 10th World Congress on Civil, Structural, and Environmental Engineering*. New Cairo, Egypt, 2025. doi: 10.11159/icsect25.182.
[12] G. M. Di Giuda, M. Locatelli, and E. Seghezzi. Natural language processing and bim in aeco sector: A state of the art. In *Proceedings of International Structural Engineering and Construction,* 7(2), 2020. doi: 10.14455/ISEC.2020.7(2).CON-22.
[13] M. Locatelli, E. Seghezzi, and G. M. Di Giuda. Exploring BIM and NLP applications: A scientometric approach. In *Proceedings of International Structural Engineering and Construction,* 8(1), 2021. doi: 10.14455/ISEC.2021.8(1).CON-12.
[14] G. G. Chowdhury. Natural language processing. *Annual Review Info Sci & Tec*, 37(1):51–89, 2003. doi: 10.1002/aris.1440370103.
[15] W. Khan, A. Daud, K. Khan, S. Muhammad, and R. Haq. Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends. *Natural Language Processing Journal*, 4:100026, 2023. doi: 10.1016/j.nlp.2023.100026.
[16] R. D. Logcher, M.-T. Wang and F. H.-S. Chen. Knowledge Processing for Construction Management Data Base. *Journal of Construction Engineering and Management*, 115(2):196–211, 1989. doi: 10.1061/(ASCE)0733-9364(1989)115:2(196).
[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention Is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010, Long Beach, California, USA. doi: 10.48550/arXiv.1706.03762.
[18] Z. Wang, Z. Chu, T. V. Doan, S. Ni, M. Yang, and W. Zhang. History, Development, and Principles of Large Language Models – An Introductory Survey. *AI Ethics*, 5:1955–1971, 2025. https://doi.org/10.1007/s43681-024-00583-7.
[19] A. Borrmann, M. König, C. Koch und J. Beetz. Die BIM-Methode im Überblick. In *Building Information Modeling*, pages 1–31, Springer Fachmedien, Wiesbaden, 2021. doi: 10.1007/978-3-658-33361-4_1.
[20] A. Gärtner, T. Bender, M. Härtig, M. Krämer, M. May, and M. Schlundt. Building Information Modeling. In *CAFM-Handbuch*, pages 197–233, Springer Fachmedien, 2025. doi: 10.1007/978-3-658-49047-8_11.
[21] H. Dawood, J. Siddle, and N. Dawood. Integrating IFC and NLP for automating change request validations. *Journal of Information Technology in Construction*, 24:540–552, 2019. doi: 10.36680/J.ITCON.2019.030.
[22] buildingSMART International Limited, Industry Foundation Classes (IFC). Online: https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes, Accessed: 06/01/2026.
[23] A. Fink. *Conducting research literature reviews: From the internet to paper*. SAGE Publications, Los Angeles, 2020.
[24] M. J. Page et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *BMJ*, 372:n71, 2021, doi: 10.1136/bmj.n71.
[25] N. J. van Eck, and L. Waltman. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523–538, 2010. doi: 10.1007/s11192-009-0146-3.
[26] B. Liu and H. Chen. BIMCoder: A Comprehensive Large Language Model Fusion Framework for Natural Language-Based BIM Information Retrieval. *Applied Sciences*, 15(14):7647, 2025. doi: 10.3390/app15147647.
[27] T. Patel and V. Patel. Data privacy in construction industry by privacy-preserving data mining (PPDM) approach. *Asian J Civ Eng*, 21(3):505–515, 2020. doi: 10.1007/s42107-020-00225-3.
[28] R. Lopez and P. E. D. Love. Design Error Costs in Construction Projects. *Journal of Construction Eng. Manage*ment, 138(5):585–593, 2012. doi: 10.1061/(ASCE)CO.1943-7862.0000454.
[29] M.D.B. in Cugurra et al. The Cross-Fertilization Between the Human-in-the-Loop Approach and the Explainable AI Techniques Toward Trustworthiness. In *Artificial Intelligence, Data and Robotics*, E. Curry et al., pages 77–104, Cham: Springer Nature Switzerland, 2026. doi: 10.1007/978-3-032-10561-5_5.