buildings

MDPI

Article

Optimising Contract Interpretations with Large Language Models: A Comparative Evaluation of a Vector Database-Powered Chatbot vs. ChatGPT

P. V. I. N. Saparamadu [1,*], Samad Sepasgozar [2], R. N. D. Guruge [3], H. S. Jayasena [1], Ali Darejeh [4], Sanee Mohammad Ebrahimzadeh [5] and B. A. I. Eranga [1]

[1] Department of Building Economics, Faculty of Architecture, University of Moratuwa, Moratuwa 10400, Sri Lanka; suranga@uom.lk (H.S.J.); isurue@uom.lk (B.A.I.E.)
[2] School of Built Environment, University of New South Wales, Sydney, NSW 2052, Australia; sepas@unsw.edu.au
[3] Department of Design Studies, Faculty of Engineering, NSBM Green University, Homagama 10200, Sri Lanka; romasha.g@nsbm.ac.lk
[4] School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia; ali.darejeh@unsw.edu.au
[5] Department of AI & Computer Engineering, Knowledge-Based Enterprise, School of Media, University of Tehran, Tehran 1411713114, Iran; sanee.sepasgozar@ut.ac.ir
* Correspondence: ishini@concolabs.com; Tel.: +94-774102909

Abstract: Frequent ambiguities in contract terms often lead to costly legal disputes and project delays in the construction industry. Large Language Models (LLMs) offer a promising solution, enhancing accuracy and reducing misinterpretations. As studies pointed out, many professionals use LLMs, such as ChatGPT, to assist with their professional tasks at a minor level, such as information retrieval from the Internet and content editing. With access to a construction regulation database, LLMs can automate contract interpretation. However, the lack of Artificial Intelligence tools tailored to industry regulations hinders their adoption in the construction sector. This research addresses the gap by developing and deploying a publicly available specialised chatbot using the ChatGPT language model. The development process includes architectural design, data preparation, vector embeddings, and model integration. The study uses qualitative and quantitative methodologies to evaluate the chatbot's role in resolving contract-related issues through standardised tests. The specialised chatbot, trained on construction-specific legal information, achieved an average score of 88%, significantly outperforming ChatGPT's 36%. The integration of a domain-specific language model promises to revolutionise construction practices through increased precision, efficiency, and innovation. These findings demonstrate the potential of optimised language models to transform construction practices.

Keywords: artificial intelligence; automated contracts; building construction; chatbot; ChatGPT; intelligent contract; large language models; natural language processing; smart contracts

Academic Editor: Yasser Mohamed

Received: 28 December 2024
Revised: 26 February 2025
Accepted: 14 March 2025
Published: 31 March 2025

Citation: Saparamadu, P.V.I.N.; Sepasgozar, S.; Guruge, R.N.D.; Jayasena, H.S.; Darejeh, A.; Ebrahimzadeh, S.M.; Eranga, B.A.I. Optimising Contract Interpretations with Large Language Models: A Comparative Evaluation of a Vector Database-Powered Chatbot vs. ChatGPT. *Buildings* **2025**, *15*, 1144. https://doi.org/10.3390/buildings15071144

1. Introduction

Contracts are fundamental to the construction industry, where projects of varying complexities unfold through collaboration among diverse stakeholders, including clients, contractors, consultants, suppliers, and engineers. These legally binding agreements define rights, responsibilities, and expectations, providing a structured framework to facilitate smooth project execution, mitigate risks, and resolve disputes [1]. Given the high financial stakes and the complexity of technical specifications, ensuring the clarity and accuracy of contract interpretation is critical for avoiding project delays and legal disputes [2]. The adoption of intelligence contracts has recently been suggested in the construction industry.

Despite the importance of precise contract interpretation, misinterpretations remain prevalent in the construction sector. According to the 2019 KPMG Global Construction Survey, 78% of construction disputes arise from contract ambiguities, leading to costly litigation and significant project delays [3]. Other studies indicate that up to 80% of construction-related legal cases involve disputes over contract terms and their interpretation [4]. These findings highlight the necessity of improving contract clarity and implementing innovative solutions to enhance accuracy in contract comprehension.

Artificial Intelligence (AI) and Natural Language Processing (NLP) have demonstrated remarkable progress in improving text interpretation across various industries, including legal and financial sectors. Large Language Models (LLMs), such as ChatGPT, have shown potential in automating document analysis, improving information retrieval, and enhancing decision-making processes [5]. However, generic LLMs lack the domain-specific training necessary to interpret complex construction contracts accurately. With its highly specific terminology and regulatory requirements, the construction industry requires AI models fine-tuned to its unique context [6].

Existing studies have explored the application of LLMs in construction-related tasks such as document generation and project scheduling [7]. While some legal sectors have adopted AI-driven chatbots for contract analysis, similar domain-specific implementations in construction remain largely unexplored. The potential for AI-enhanced contract interpretation in this sector presents an opportunity to address existing challenges in contract management and dispute resolution [8].

This study aims to bridge this gap by developing a domain-specific chatbot leveraging LLMs for contract interpretation in the construction industry. Domain-specific chatbot development has received more attention in recent years. The research investigates the effectiveness of a customised AI model trained on construction-specific legal data and compares its performance with a general-purpose LLM, ChatGPT. The primary objectives of this study are as follows:

To assess the capabilities of five prominent LLMs in interpreting construction contracts.

To investigate the adoption of LLMs in the construction sector for contract-related applications.

To develop an optimised AI model tailored for construction contract interpretation.

To evaluate the accuracy of the domain-specific chatbot compared to the baseline ChatGPT model.

By integrating AI into contract management, this research seeks to enhance precision, reduce legal disputes, and promote efficiency in construction project execution. The findings will contribute to the growing body of knowledge on AI applications in the construction industry and provide valuable insights for professionals seeking to adopt AI-driven solutions for contract analysis and risk mitigation.

2. Methodology

The paper presents details of a tool development, including the experiment and tests carried out, to discuss the accuracy and reliability of the proposed model. It can also be said that the study adopts a pragmatist philosophy, allowing for a comprehensive research problem analysis. Pragmatism is a philosophical tradition that evaluates ideas and theories based on their practical consequences and real-world applications rather than solely on abstract principles or inherent truths [9]. It suggests that the meaning of concepts is best understood through their effects and usefulness in solving problems, emphasising action, experimentation, and the iterative refinement of ideas [9]. The pragmatic approach supports flexible and adaptive research strategies that prioritise using qualitative and quantitative methods to address real-world problems.

In the context of this research, the primary goal is to design, develop, and assess an LLM-powered chatbot for contract use. Therefore, the pragmatic approach aligns with evaluating the real-world impact of integrating LLMs into contract interpretation. Furthermore, one of the core tenets of pragmatics is that language must be understood in its specific use-case scenario. Since contract language in construction can be ambiguous and context-dependent, a pragmatic philosophy underpins efforts to tailor language models to handle such contextual nuances accurately. Finally, the pragmatic approach bridges theoretical models of technologies such as LLMs and their practical implementation in real-world scenarios. Given the complexities of construction law and the frequent misinterpretations that can lead to legal disputes, it provides a solid foundation for justifying why a domain-specific adaptation of ChatGPT is necessary.

The study employs a mixed-method approach, combining both qualitative and quantitative methods to comprehensively investigate the development and performance evaluation of chatbots capable of answering contractual questions in the construction industry. To do so, a custom-built chatbot was developed to ascertain the answers to questions about contract interpretation that had been generated. Figure 1 elaborates on the methodology used.

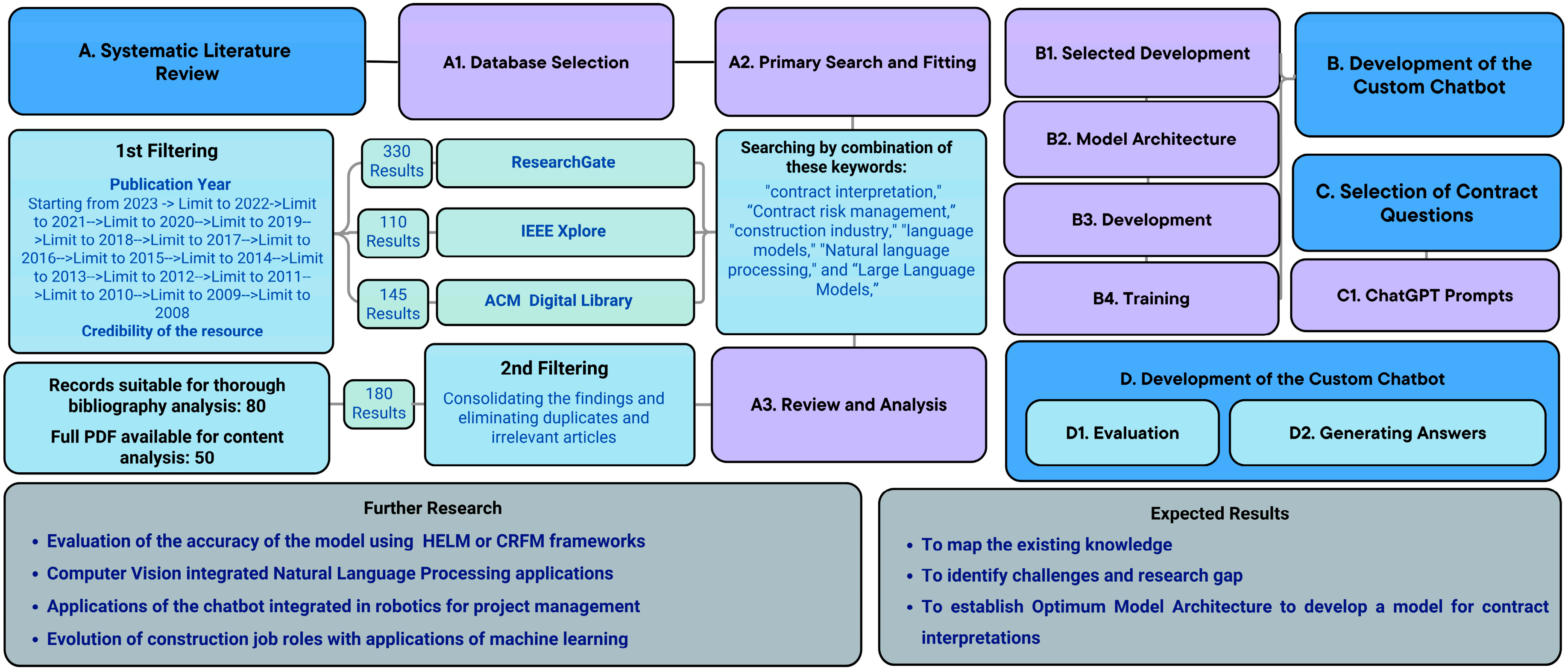

A. Systematic Literature Review
A1. Database Selection
A2. Primary Search and Fitting
B1. Selected Development
B. Development of the Custom Chatbot
1st Filtering
Publication Year
Starting from 2023 -> Limit to 2022->Limit to 2021-->Limit to 2020-->Limit to 2019-->Limit to 2018-->Limit to 2017-->Limit to 2016-->Limit to 2015-->Limit to 2014-->Limit to 2013-->Limit to 2012-->Limit to 2011-->Limit to 2010-->Limit to 2009-->Limit to 2008
Credibility of the resource
330 Results
ResearchGate
110 Results
IEEE Xplore
145 Results
ACM Digital Library
Searching by combination of these keywords:
"contract interpretation," “Contract risk management,” "construction industry," "language models," "Natural language processing," and “Large Language Models,”
B2. Model Architecture
B3. Development
B4. Training
C. Selection of Contract Questions
C1. ChatGPT Prompts
Records suitable for thorough bibliography analysis: 80
Full PDF available for content analysis: 50
180 Results
2nd Filtering
Consolidating the findings and eliminating duplicates and irrelevant articles
A3. Review and Analysis
D. Development of the Custom Chatbot
D1. Evaluation
D2. Generating Answers
Further Research
• Evaluation of the accuracy of the model using HELM or CRFM frameworks
• Computer Vision integrated Natural Language Processing applications
• Applications of the chatbot integrated in robotics for project management
• Evolution of construction job roles with applications of machine learning
Expected Results
• To map the existing knowledge
• To identify challenges and research gap
• To establish Optimum Model Architecture to develop a model for contract interpretations

Figure 1. Research flow diagram (Source: Developed by Authors).

2.1. Qualitative Methods

The methodology includes an extensive literature review as a foundational component. This literature review systematically examines existing scholarly works related to chatbots in the context of the construction industry and their applicability to contractual inquiries. A comprehensive literature review is conducted on chatbots and their applications within the construction industry. This review aims to identify key trends, best practices, and gaps in the current research landscape.

Literature Review Method

Step 1: Database Selection

Choosing relevant sources for the present study, centred on the "Optimal Language Model for Contract Interpretation in the Construction Industry", involved the selection of three reputable and comprehensive databases: ResearchGate, IEEE Xplore, Scopus, and ACM Digital Library. These databases were chosen due to their combined capacity to offer extensive coverage of relevant articles related to the research topic. Utilising multiple databases minimises the risk of overlooking any important papers in the field.

Step 2: Primary Search

To initiate the search, a set of initial keywords was utilised to guide the accuracy of search results. The following keywords were chosen as primary factors for controlling the search results: "interpretation of contracts", "management of contract risks", "construction sector", "language models", "NLP", and "Large Language Models". A thorough examination of recent publications within the field was conducted to guarantee the inclusion of all pertinent keywords and to identify and compile key subtopics and keywords. These keywords were chosen to ensure the inclusion of all relevant existing sources in the database, leaving no gaps or potential exclusions.

In order to obtain an overview of the current themes in this field, a wider range of keywords was also used for searching within Scopus. The search string combined three sets of keywords focusing on artificial intelligence, construction, and contract challenges. This is called Search 2, and the main string was as follows:

(TITLE-ABS-KEY("natural language processing" OR "NLP" OR "AI" OR "artificial intelligence" OR "language model" OR "language technology" OR "large language model" OR "foundation model" OR "generative AI" OR "transformer model" OR bot OR chatbot OR ChatGPT) AND TITLE-ABS-KEY("construction industry" OR "building construction" OR "project management" OR "construction project") AND TITLE-ABS-KEY(contract OR risk OR dispute OR claim OR interpretation OR ambiguity)).

Figure 2 presents a network visualisation based on bibliographic data extracted from the Scopus database named Search 2. The data spans research from the year of the first publication to 2025, capturing key trends, connections, and collaborations in the field. The network highlights relationship keyword co-occurrence, offering insights into emerging topics and the evolution of research themes, where AI and project management are the most frequent keywords.

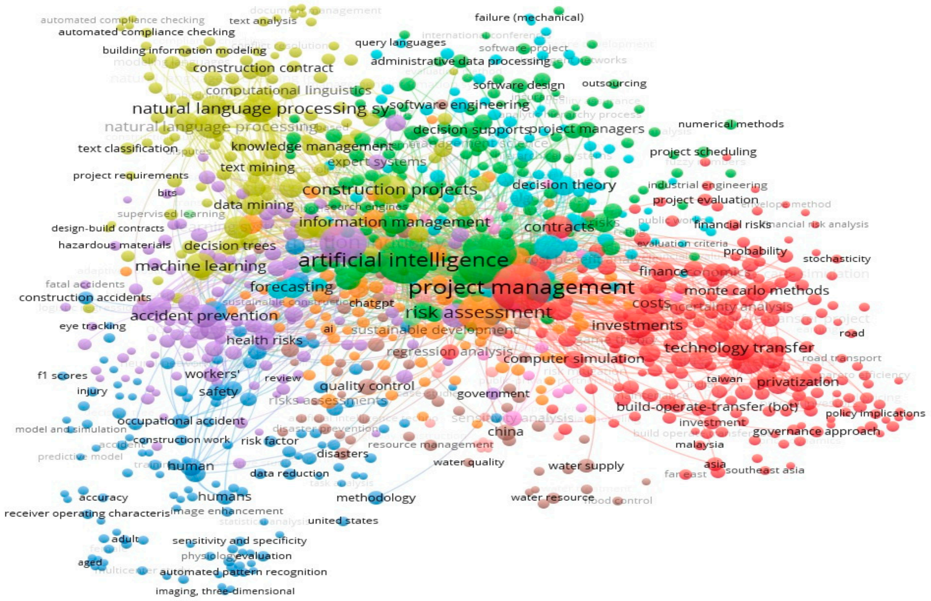
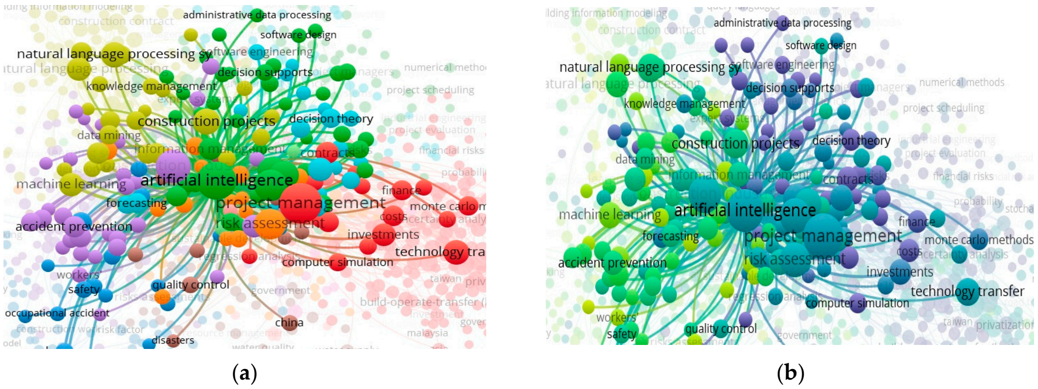
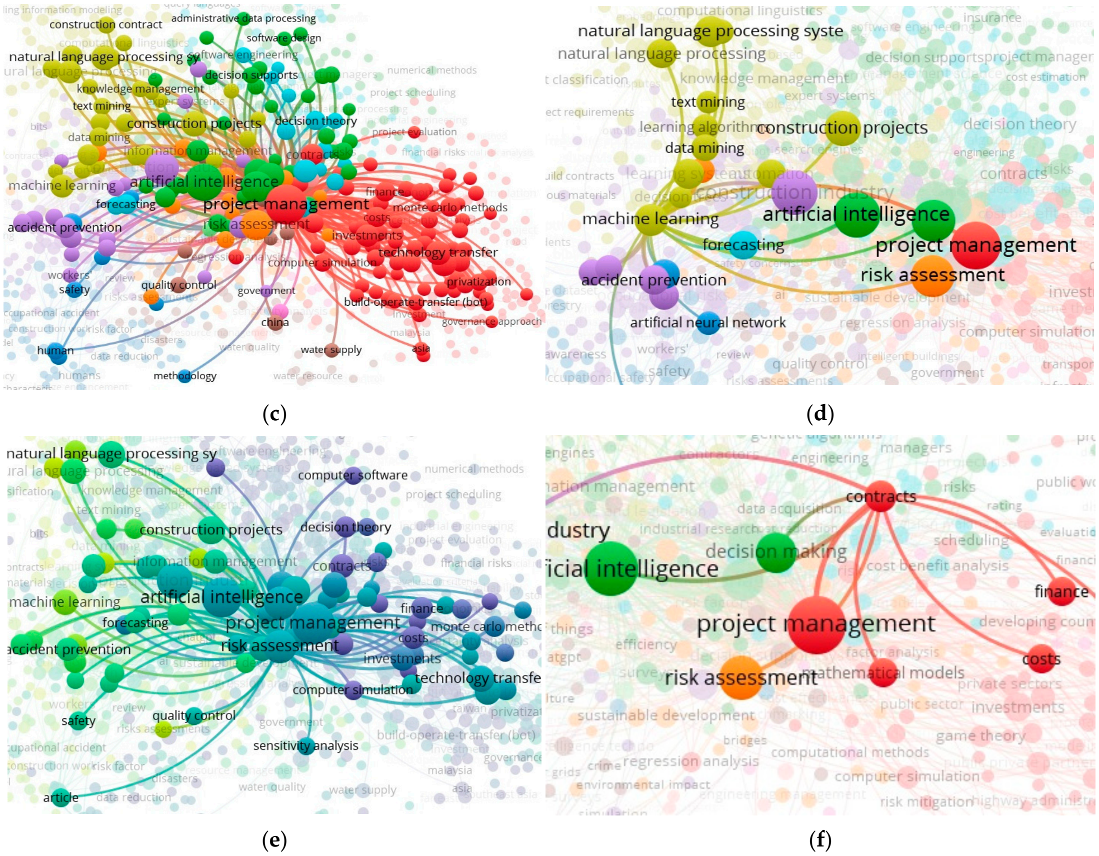


Figure 2. Bibliographic data extracted from the Scopus database, covering research from its inception in 1985 to 2025.

Figure 2 shows the network distribution of studies in the literature and reveals that the key focus areas are AI, project management, and risk management. In contrast, fewer studies address other themes, such as contract management and compliance issues using AI. The figure also highlights emerging topics such as NLP systems, computational linguistics, construction contracts, and knowledge management. These areas have the potential to be further enhanced through the application of AI and intelligent systems.

Figure 3 presents a network visualisation segmented by year, using distinct colour codes to illustrate the evolution of research connections from inception to 2025. The visualisations focus on various themes within the literature, including AI (a), project management (b), risk assessment (c), machine learning (d), expanded risk assessment (e), and contracts (f). The figure reveals that AI, project management, and risk assessment (visualisations a, b, and c) demonstrate highly intensive networks with dense keyword concentrations, reflecting significant discussions and well-established research areas. Most other research themes appear closely related to these central topics, indicating their dominant role in the field.





(c)

(d)

(e)

(f)

Figure 3. Network visualisation centred on AI, segmented by year through distinct colour codes, illustrating the evolution of research connections from inception to 2025. (**a**) Visualisation centred on AI; (**b**) visualisation centred on project management; (**c**) visualisation is centred on risk assessment and segmented; (**d**) visualisation is centred on machine learning; (**e**) visualisation is centred on risk assessment; and (**f**) visualisation is centred on contracts.

In contrast, machine learning, risk assessment, and contracts (visualisations Figure 3d–f) emerge as relatively less developed themes, suggesting they are still evolving based on the advancement of technologies in the field. This highlights considerable gaps for further exploration, particularly in applying machine learning, AI systems, virtual assistants, and bots to areas like contract management and risk assessment, specifically focusing on contractual considerations and compliance. These networks offer valuable insights into the gaps within the existing body of research and underscore areas where more in-depth investigations and contributions are needed to advance the field.

Step 3: Bibliography and Content Analysis

From all databases, 180 results were checked and filtered manually to exclude duplicates, non-English papers, full PDF availability, and less relevant articles that the previous steps had not excluded. This resulted in 50 articles being analysed through clustering and in-depth review.

2.2. Quantitative Methods

Quantitative methods will be employed to evaluate the performance of various chatbots designed to answer contractual questions. This quantitative phase is based on a project conducted at the University of Moratuwa. For that project, a standardised set of test questions related to contractual matters in the construction industry was created. These questions serve as the basis for evaluating chatbot responses. Subsequently, answers to these pre-determined questions are generated through the custom-made chatbot and ChatGPT. Data on the accuracy and efficiency of chatbot responses will then be systematically collected and recorded. Then, these data will be evaluated through the cosine distance evaluation method, a reputed AI accuracy measurement metric (Figure 4).

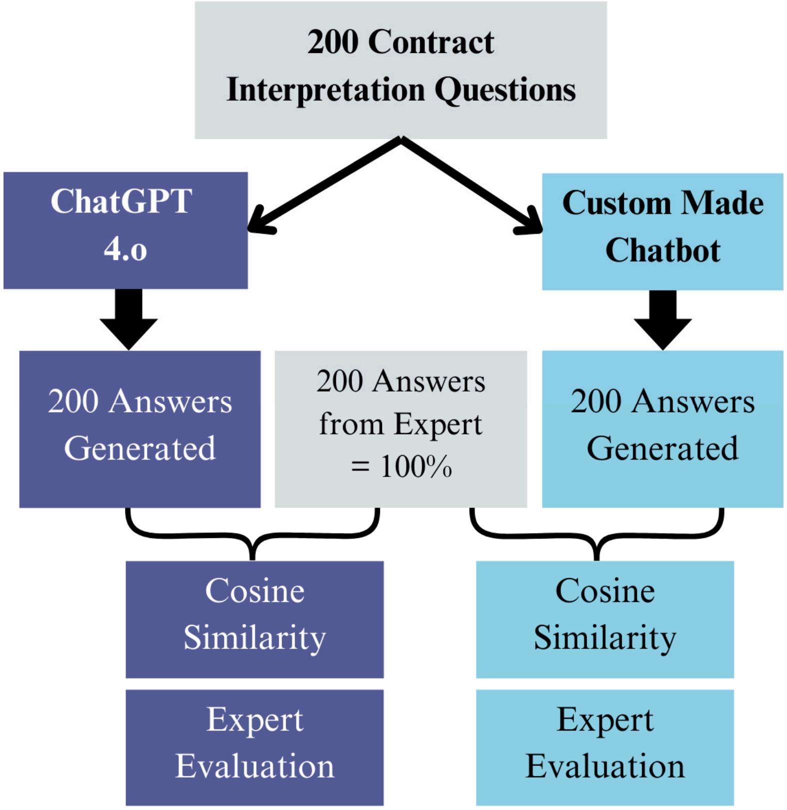


Figure 4. Methodology of the Research Analysis (Source: developed by authors).

3. Analysis of the Literature Relevant to the Study

3.1. Contract Interpretation and Management in the Building Sector

A research study aimed at identifying strategies to address construction contracts issues during the construction process has highlighted several reasons why problems related to contracts in the built environment often become complex [10]. The findings have delineated these causes as follows: unclear definitions regarding the rights and responsibilities of the contract parties, imprecise outlining of how penalties are calculated for missing deadlines, insufficiently detailed specifications of tasks and milestones, absence of provisions regulating changes to project documentation during construction, excessive contractual penalties primarily imposed on the contractor, and a lack of provisions governing procedures for carrying out additional and replacement tasks, along with their corresponding resolutions [3,4,11].

However, to manage these conflicts appropriately, the parties involved should utilise strategies to reduce uncertainties that align with different project phases [12]. These strategies include establishing contingency plans, ensuring construction guarantees, presenting claims for time extensions, providing payment guarantees, and incorporating clauses for retention and cost adjustments [13]. Conversely, to mitigate contract disputes, construction professionals should also understand the strategies that must be applied before and more effectively and proactively throughout the project implementation. Therefore, when these strategies are executed proficiently in tandem, they will significantly contribute to the success of a construction project and result in a noticeable reduction in contractual disputes [1].

Another study suggests that when interpreting contracts, it is important not to overplay the criterion of "commercial common sense". The study further argues that while commercial common sense can be useful in interpreting contracts, it should not be relied upon too heavily as it can lead to incorrect interpretations. Instead, a more literal approach to contractual interpretation should be taken when dealing with construction contracts [14].

3.2. NLP and Construction Decision-Making

In the dynamic realm of construction, where the projected global expenditure is set to reach $17.5 trillion by 2030, China, the US, and India are at the forefront, collectively driving 57% of worldwide growth [15]**.** Considering this, finding innovative approaches to tackle the industry's inherent challenges becomes paramount. Intricate communication networks, complex documentation, and a diverse array of stakeholders characterise the landscape of construction projects. This demands a level of sophistication that goes beyond conventional methodologies. This is precisely where NLP can prove invaluable [6]. NLP is a conduit for computers to comprehend and effectively utilise human language, potentially revolutionising how we oversee construction projects. Since the construction field relies heavily on the exchange of textual documents, NLP can offer a solution to surmount the sector's document-centric nature [16].

NLP techniques, focused on bridging the gap between human language and computer understanding, have showcased their transformative potential across various domains, with an impressive 89% accuracy achieved in sentiment analysis studies [6,17]**.** Smith and Johnson's seminal work illuminates how NLP techniques empower the construction industry to harness the power of language for efficient and effective management [18]. By enabling the analysis of textual data, NLP techniques offer a gateway to valuable insights. One notable application is sentiment analysis, which involves the automated assessment of stakeholders' sentiments and attitudes. This analytical prowess has the capacity to enhance stakeholder communication by providing a real-time assessment of project perceptions, enabling project managers to tailor their strategies accordingly [19].

Another significant application lies in the realm of text summarisation, a process by which voluminous project documentation can be distilled into concise yet comprehensive summaries [20]. This possesses far-reaching implications for streamlining project reporting, documentation, and knowledge dissemination. Through text summarisation, the intricate details within technical reports, design specifications, and communication logs can be distilled into easily digestible formats, fostering efficient decision-making and facilitating collaboration among project teams [21].

Furthermore, using NLP techniques extends to entity recognition, a critical process for identifying and categorising specific elements within textual data. In the context of construction, entity recognition can aid in identifying key project components, such as materials, equipment, project phases, and stakeholders [22]. This enhances the accuracy of information retrieval and lays the groundwork for sophisticated analyses and predictive modelling. Integrating NLP techniques into the construction industry signifies a profound paradigm shift underpinned by compelling reasons and substantial evidence illuminating its transformative potential. By leveraging the power of NLP, professionals in the construction industry can open doors to a realm of new possibilities. This can potentially bring revolutionary changes to project management practices in many ways [23].

Initially, the potential of NLP techniques to examine text data and conduct sentiment analysis provides construction stakeholders with immediate insights into project perceptions [18]. This instant evaluation of sentiments allows project managers to promptly tackle concerns and adapt communication strategies as needed, promoting better stakeholder engagement and collaboration. Furthermore, sentiment analysis driven by NLP supplements conventional feedback channels by offering a data-driven, impartial viewpoint on stakeholder sentiments, thereby enriching the overall decision-making process [24].

Furthermore, applying NLP techniques in text summarisation transforms the landscape of project documentation [25]. NLP-driven text summarisation expedites decision-making processes by condensing voluminous technical reports, design specifications, and communication logs into concise yet comprehensive summaries. It facilitates effective knowledge dissemination among project teams. As affirmed by studies, this efficiency enhancement reduces administrative overhead, accelerates project timelines, and minimises the risk of critical information being overlooked [26].

3.3. Types of Available LLMs

In the ever-advancing landscape of modern technology, new iterations of large LLMs are frequently unveiled, each designed to tackle greater complexities than their predecessors. The following paragraphs evaluate the most prominent LLMs. In recent studies by Rane et al. [27], Li and Li [28], and He et al. [29], conducted for the construction industry, the authors cumulatively remarked ChatGPT 4.0, ChatGPT-3.5, BERT, RoBERTa, and Transformer-XL as being the most prominently used LLMs.

OpenAI introduced ChatGPT as an AI model tailored for interactive conversations. This model is closely related to InstructGPT, designed to comprehend prompts and provide detailed responses [30]. ChatGPT originated from the GPT-3.5 series and completed its training in early 2022 [31]. ChatGPT-4.0 builds upon the capabilities of its predecessors with improved contextual understanding, broader knowledge, and enhanced conversational abilities. It leverages deep learning techniques to generate human-like text, making it a powerful tool for various applications, from customer service to creative writing.

BERT, an acronym for Bidirectional Encoder Representations from Transformers, constitutes a family of language models introduced in 2018 by a team of researchers at Google [32]. The foundational research paper outlining BERT is titled “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” [32]. The primary objective of BERT is to pre-train extensive bidirectional representations from unlabelled text, accomplishing this by concurrently considering both the preceding and succeeding contexts across all layers [33]. Notably, achieving such advancements does not necessitate substantial modifications to task-specific architectures.

XLNet acts as a type of language model called a Transformer. It combines two strong techniques: one that guesses words based on what happened earlier and another that looks at the whole sentence while avoiding problems [34]. Instead of being relieved to a fixed order, it looks at all the different ways to break down a sentence and predicts words from that. This helps it use words from both sides. Each part of a sentence learns from the rest, taking in information from both directions.

The Transformer-XL architecture, introduced in 2019 by researchers from Carnegie Mellon University and Google, is an advancement of the Transformer framework tailored to enhance the model’s capability to capture distant dependencies within sequential data [35]. This innovation introduces two crucial techniques: a segment-level recurrence mechanism and a relative positional encoding scheme [16]. In the training process, the representations computed for preceding segments are cached and preserved for extended context when the model processes subsequent new segments. This integration facilitates the flow of contextual information across segment boundaries, resulting in a remarkable increase in the potential dependency length by a factor of N, representing the depth of the network. To prove its efficacy, Transformer-XL has demonstrated exceptional performance across various language modelling tasks, encompassing character-level and word-level language modelling [36]. The availability of Transformer-XL on platforms like Hugging Face has made it a popular choice among researchers and developers engaged in NLP endeavours.

Table 1 features ChatGPT-3.5, BERT, RoBERTa, Transformer-XL, and GPT-4 comparisons based on functionalities, limitations, cost, and accuracy.

Table 1. Language model tools**.**

Language Modelling Tool and References	Characteristics	Limitations	Accuracy	Cost
ChatGPT (GPT4.o) [37,38]	Improved version of ChatGPT-3.5, optimised for more recent data and larger batches	Potential for biassed outputs	High for conversational tasks but may lack domain-specific precision	High due to the large model size
ChatGPT-3.5 [37,38]	Generates human-like text, good for conversational AI, creative writing, and information retrieval	Limited by training data cut-off (2021), the potential for biassed outputs	High for conversational tasks but may lack recent information	Moderate, depending on the use
BERT [38,39]	Excels in NLP tasks like text classification, named entity recognition, and question-answering	Not optimised for text generation, limited by fixed input size	High accuracy in downstream tasks (e.g., SQuAD)	Moderate, depending on application and fine-tuning needs
RoBERTa [34,39]	Improved version of BERT, optimised for more extensive training data and larger batches	Still lacks the ability for text generation and requires significant computational resources	Higher accuracy than BERT on most NLP benchmarks	Higher than BERT due to extended training costs
Transformer XL [16,38]	Handles long-range dependencies better than traditional transformers, useful for tasks requiring long context	Complexity in training, potential instability in very long sequences	High accuracy in tasks requiring longer contextual understanding	High due to complex architecture and resource needs

3.4. Adopting ChatGPT to Construction

Building upon the foundational principles of NLP, the emergence of LLMs represents a groundbreaking advancement in construction technology, with ChatGPT standing out as a prime example of this innovation [40]. These models have revolutionised how textual data are processed, analysed, and generated, offering unparalleled linguistic comprehension and contextual reasoning capabilities. As recent studies have demonstrated [41–43], LLMs are not only adept at understanding complex linguistic structures but also exhibit remarkable proficiency in identifying nuanced contextual variations and generating semantically coherent, contextually appropriate responses [44]. These capabilities make them highly adaptable for domain-specific applications, including the construction industry, where precise communication, documentation, and project management are paramount.

One of the critical advantages of ChatGPT and similar models lies in their ability to be fine-tuned and tailored to highly specialised domains. Through a rigorous process of domain-specific training, ChatGPT can be adapted to the intricate linguistic and technical demands of construction-related discourse [8]. This customization allows the model to grasp industry-specific terminology, understand procedural intricacies, and interpret project requirements with a level of precision previously unattainable through conventional automation tools. As a result, ChatGPT emerges as a powerful asset in augmenting efficiency, accuracy, and decision-making processes across various facets of construction technology.

Figure 5 showcases research conducted in recent years concerning adapting ChatGPT to construction. It shows that a constrained group of authors is co-authoring their work, indicating that the scholars' network is nearly disconnected.

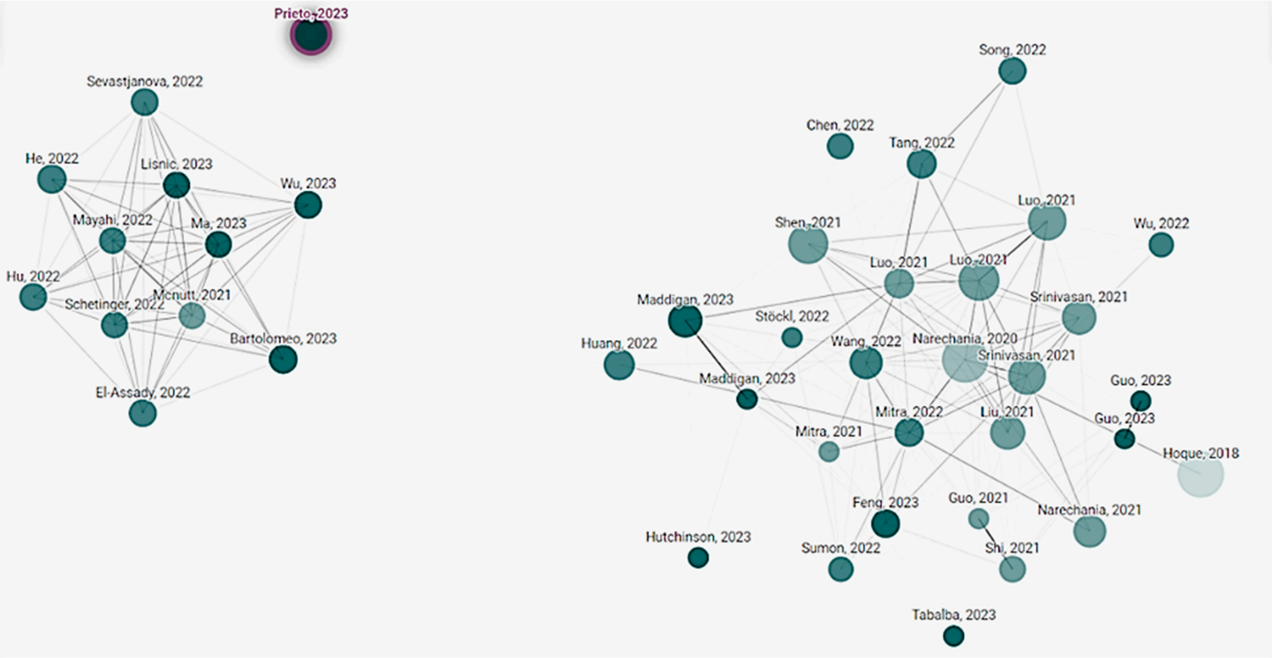


Figure 5. A network of studies referring to ChatGPT applications in construction.

3.5. Application of ChatGPT in Enhancing Project Management and Reporting Efficiency

Among the most transformative applications of ChatGPT in construction is its capacity to automate and optimise project reporting—a traditionally time-intensive and error-prone task. Studies indicate that leveraging ChatGPT for project reporting can reduce the time required by up to 70% while improving report accuracy by 85% [45]. The model's advanced natural language capabilities enable it to process vast quantities of project-related data, extract relevant insights, and structure them into clear, coherent, and actionable reports. By automating this process, project managers and stakeholders gain access to timely, high-quality documentation that enhances transparency, accountability, and operational efficiency [46].

Furthermore, ChatGPT's integration into project reporting systems mitigates human-induced inconsistencies and enhances the standardisation of reports. Traditional reporting methods often suffer from variations in phrasing, inconsistent formatting, and occasional omissions due to human error. ChatGPT ensures uniformity in terminology, structure, and presentation, thus, facilitating more reliable and comparable documentation over the lifecycle of a project. This standardisation proves particularly beneficial in large-scale construction projects involving multiple stakeholders, where clear and consistent communication is essential for effective coordination.

3.6. Application of ChatGPT in Construction Workflows and Decision-Making

Beyond project reporting, the integration of ChatGPT into construction workflows extends to multiple operational domains. The model's ability to analyse historical project data, identify trends, and generate predictive insights enables proactive decision-making. For instance, ChatGPT can assist in risk assessment by analysing past project reports, flagging recurring issues, and suggesting preventive measures [27,44]. Additionally, its conversational AI capabilities facilitate interactive assistance for field engineers, allowing them to query technical documentation, regulatory guidelines, and best practices in real-time.

Moreover, ChatGPT's automation potential contributes to cost reduction by minimising the manual effort required for documentation and administrative tasks. Companies can achieve higher productivity levels and optimise resource utilisation by reallocating human resources to more strategic and creative functions. This shift enhances efficiency and fosters innovation by enabling construction professionals to focus on problem-solving and strategic planning rather than routine paperwork [47].

3.7. Application of ChatGPT in Contract Interpretation

Addressing legal and contractual concerns is paramount in the construction industry, ensuring project success and pre-empting disputes [48]. Contract interpretation is an essential aspect of the construction industry, as it helps ensure that the contract terms are understood and followed by all parties involved. However, the contract interpretation is a legal matter in the province of judges and arbitrators, not the parties to the contracts themselves. Manual contract interpretation is inherently prone to subjectivity, human errors, and time constraints, often leading to misinterpretations that can escalate into costly legal disputes [49].

Recent advancements in the field of NLP offer a promising avenue for streamlining contract management practices. Hassan, Fahad Ul, Le Tuyen, and Le Chau [50] presented an innovative NLP framework tailored to digitalise nonquantitative natural language requirements in construction contracts. This framework encompasses a comprehensive array of semantic and syntactic rules designed to extract crucial contractual information, including stakeholders, actions, objects, constraints, tasks, and obligations. The efficacy of the proposed model was designed for design-build highway contract provisions, and it had notable precision and recall rates of over 93% and 87%, respectively.

Studies have been conducted to analyse the utilisation of ChatGPT for contract drafting and analysis. These studies have shown that contract professionals must focus on acquiring the skill of accurately priming and prompting the model to generate more precise and useful outcomes [51]. This practice also helps minimise the risk of human errors, thereby ensuring the comprehensive review and analysis of legal documents [44].

Further analysis based on Search 2 resulted in 1087 outcomes. The search was conducted in December 2024. The analysis of the outcome is shown in Figure 6.

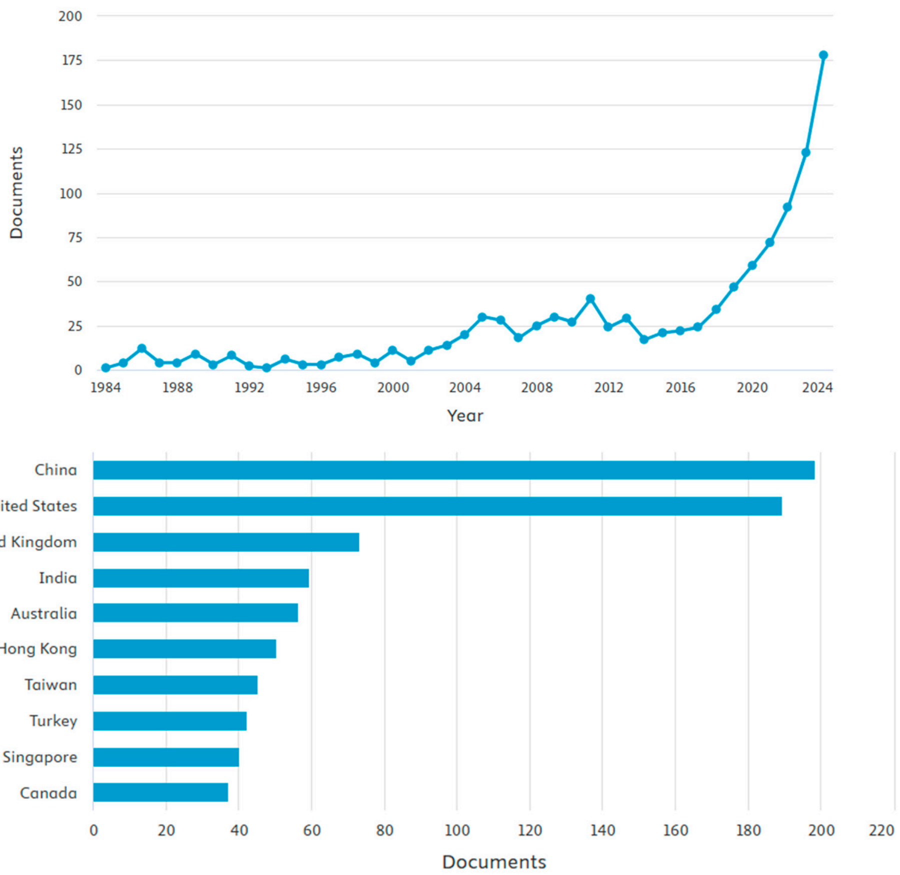


Figure 6. The trends of papers published in the field and the contribution of countries based on the number of publications.

4. Development of the Chatbot

4.1. Background to the Chatbot Development

Using and managing unstructured data is one of the most important difficulties construction organisations confront today. These data include various information, from project reports and drawings to communications and contracts. Without the proper equipment and technology, this knowledge is frequently left untapped and hidden in inaccessible documents and files. The architecture, workflow diagram, training, and evaluation of the chatbot model are as follows.

4.2. Feasible Development Models

There are two main methods for creating chatbots that use NLP. The first approach addresses contractual conundrums by employing ChatGPT's Version 4.0. However, this strategy faces significant difficulties because of a lack of data trained expressly for building and a potential shortage of domain knowledge inside the model.

ChatGPT's comprehension, as a problem-solving tool, depends on the availability of relevant data to train the model successfully. While ChatGPT is adept at producing writing that appears human, its ability mostly depends on the thoroughness and applicability of the data utilised during its training phase. Since the newest ChatGPT model could access almost all data on the World Wide Web, the model would use blogs, articles, and web pages about construction contracts to answer questions. However, ChatGPT's lack of specific topic experience might lead to mediocre replies that miss the nuances of contractual issues, particularly in the construction sector.

The second strategy comprises creating a particular model for interpreting contracts, especially in the construction industry. Although promising in adapting solutions to particular sector demands, this method has its difficulties. The extensive time and resources needed for creating, instructing, and improving such a model are one of the main obstacles. Due to problems with model complexity, integration challenges, and rapid change within the sector, past attempts described in the literature review to implement custom-built contract interpretation models within the construction industry have frequently failed.

4.3. Developing a Customised Tool

Language models offer a transformative solution for processing or making conversations with various documents, such as project reports, contracts, building information models, and emails, in the construction industry context. Their utility extends to tasks such as information extraction, content summarisation, and even responding to queries related to document contents. For example, the chatbot can be modelled for inquiries like, “What will happen if the agreement was not signed?” and questions like “Identify risks associated with this project?”.

However, to enable such a function, a mechanism needs to be developed to point to language models to navigate and extract relevant information. The mechanism lies in integrating vector databases and embeddings [52].

Vector databases, designed for handling vector data, play a pivotal role in this process. In language models, vector data takes the form of embeddings. These embeddings represent words or phrases as vectors, essentially points in a multi-dimensional space. Consequently, language models can discern word or phrase meanings based on relative positions.

Developing a customised tool should combine ChatGPT with a large vector database containing information about the building. This plan intends to give ChatGPT-specific details to increase its effectiveness. The vector database would be a helpful source of knowledge on the industry, allowing ChatGPT to access a wide range of contextually pertinent information while replying to contractual inquiries. This method may be able to fill the gap between general NLP models and industry-specific needs, providing a solid answer to contractual problems in the building industry.

When a language model processes a document, it converts the text into embeddings. These embeddings are subsequently stored within a vector database, facilitating efficient data search and comparison. To appreciate this technology’s efficacy, comparing it to conventional keyword-based searches is informative. In typical search engines, keyword searches rely on exact word matches within documents, neglecting contextual nuances and potential synonyms. In contrast, language models with vector similarity search capabilities transcend these limitations [52]. They interpret the entirety of a query, seeking documents or document segments with semantically akin meanings. Consequently, even if a document lacks the precise words in a query, the model can still locate it if its overall purpose aligns. This revolutionises the process of searching unstructured data, enabling nuanced and precise searches that reveal insights that conventional keyword searches might overlook.

Programming Languages Used

The following languages are used for the development of the chatbot:

- Python Language—Because of its compatibility with Lnagchain, data science models usually use Python 13.10.2 Version.
- Javascript—React Framework is the most often used technology to develop web applications. It is used to create a simple frontend.

4.4. Data Preparation

The Vector database is developed employing standardised forms of contract documents as its foundation. This structured approach entails extracting pivotal contractual elements, whereby the clause number is designated as the clause_id, the clause heading as the clause_name, and the substantive content within the clause as the responsibility. These data are subsequently organised into a tabular format, succinctly delineated under the categorical headings of clause_id, clause_name, and responsibility. Following meticulous collation, this dataset, enriched with the 15 most prevalent standard contract forms (refer to Figure 7), is methodically archived in the Comma Separated Values (CSV) format. The diagram below represents the standard forms of contract that have been used to develop the database.

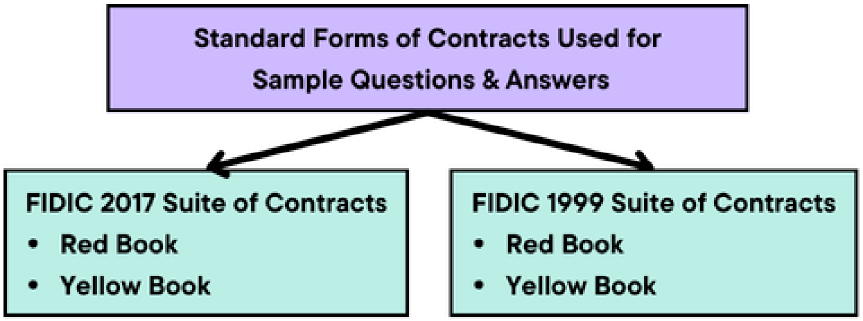

Standard Forms of Contracts Used for Sample Questions & Answers
FIDIC 2017 Suite of Contracts
• Red Book
• Yellow Book
FIDIC 1999 Suite of Contracts
• Red Book
• Yellow Book

Figure 7. Standard forms of contract that were used in the database (source: developed by authors).

4.5. Model Architecture

Figure 8 shows the available architecture models when establishing a collaboration of a vector database and a language model.

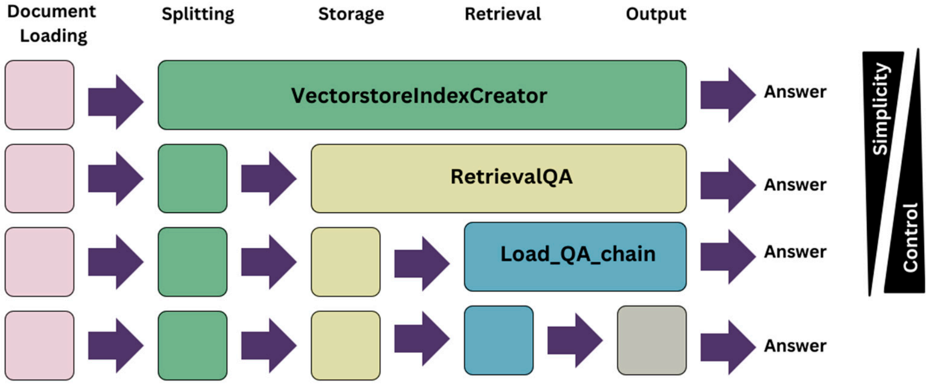


Figure 8. Language model and vector database synergising models (source: developed by authors).

Due to the significance of contractual interpretation in the construction sector, the final approach was chosen since it provides greater control, and simplicity was not a factor of consideration. The architecture of the chatbot is described as follows (refer to Figure 9).

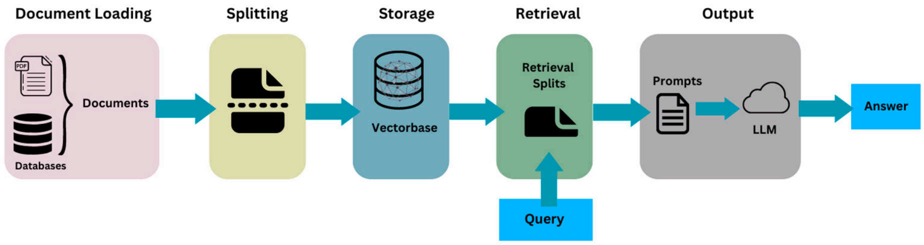


Figure 9. Architecture of the chatbot (source: developed by authors).

To achieve this architecture, the following technologies have been used:

- Open AI (Chat GPT-3)
- Langchain Library
- Chroma Database as the Vector Store

4.6. Libraries and Technologies Used

4.6.1. ChatGPT

The GPT-3.5 language model, a transformer-based neural network with many parameters, is used to develop a construction chatbot for contract interpretation. Since the introduction of the transformer design by Vaswani et al. [53], it has been a popular option for problems involving NLP Transformer architecture, which consists of several layers, each of which serves a distinct purpose [53]. The essential component of the transformer is the attention mechanism, which enables the model to evaluate the relative relevance of various words in a phrase. The formula below describes the attention mechanism.

$$Attention\ (Q, K, V) = softmax(\frac{Qk^T}{\sqrt{d_k}})\mathrm{V} \tag{1}$$

Here, Q, K, and V are matrices representing the query, key, and value, respectively. The attention mechanism calculates a weighted sum of the values based on the similarity between the query and key.

The architecture of the construction chatbot for contract interpretation is similar to that of GPT-3.5. However, it has been fine-tuned on a large corpus of legal text related to construction contracts. This fine-tuning process involves updating the model parameters to better understand and generate text specific to construction contracts. By leveraging the transformer architecture and its training on a substantial corpus of construction contract text, this chatbot excels in comprehending and generating content relevant to the construction industry’s legal aspects.

4.6.2. Langchain Library

LangChain is a platform for constructing language model-powered apps. It offers developers modular abstractions for the components required to deal with language models and collections of implementations for each abstraction [54]. The framework is intended to facilitate building robust and distinctive applications that invoke a language model via an API, connect it to other data sources, and allow it to interact with its environment.

4.6.3. Chroma and Word Embedding

Chroma is an AI-native open-source database that is used to store word embeddings.

4.6.4. Word Embedding

A word vector is an attempt to convey the meaning of a word quantitatively. A computer analyses the text and determines how frequently words appear next to one another. First, it is necessary to analyse why word vectors are seen as an advancement over standard word representations [55]. This was used to develop the vector database.

4.7. Workflow Diagram

The diagram below illustrates the model's workflow. It provides the end-to-end workings of the diagram from when the user inputs their question to when they receive the response along with the utilised sources. In addition, the data preparation stage workflow is included. The steps in the figure will be explained in the following sections, referring to the steps highlighted in Figure 10.

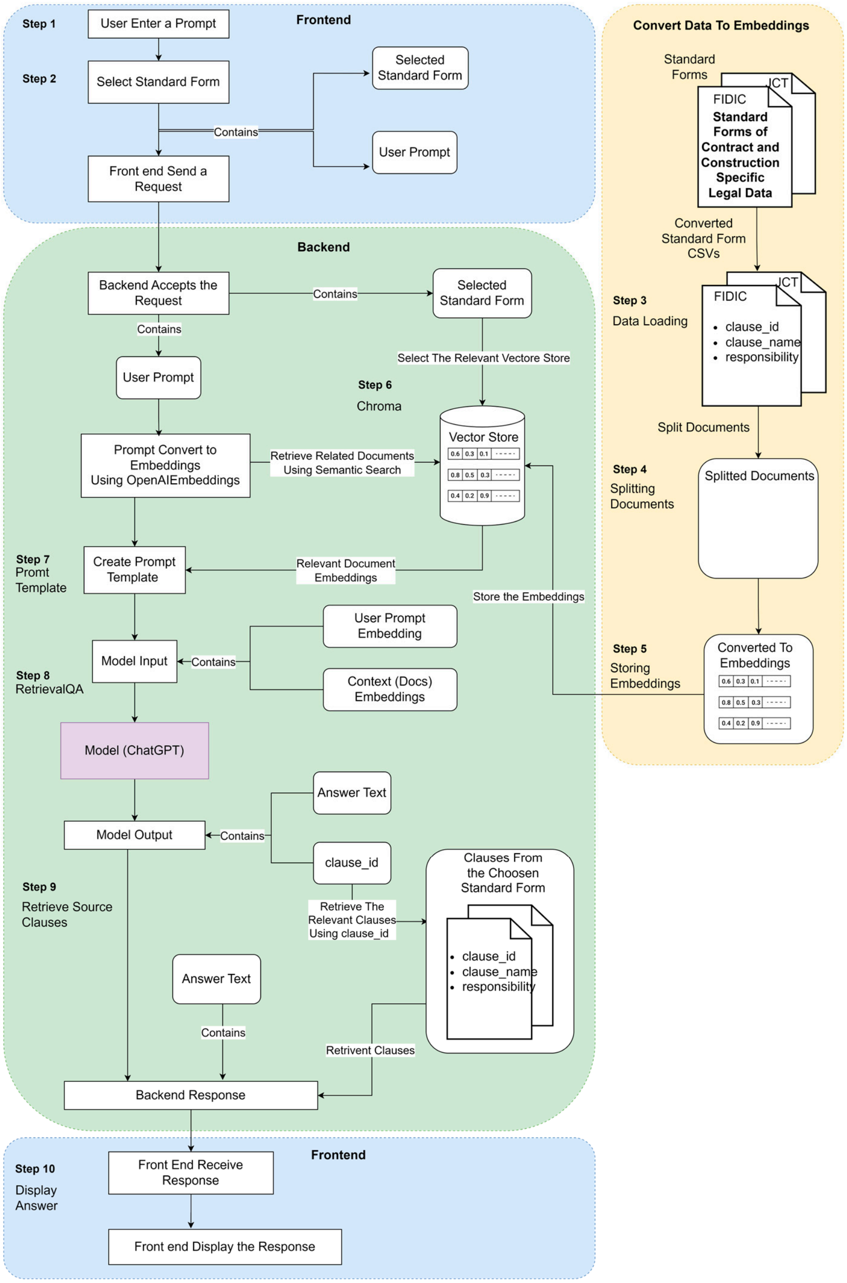


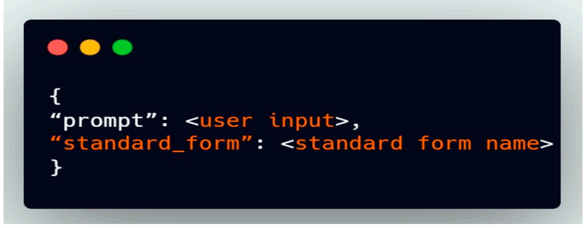
Figure 10. Workflow diagram of the chatbot (source: developed by authors).

4.7.1. User Input (Step 1)

The user interface resembles that of popular messaging apps. There are Chat Bubbles for displaying the bot’s response and the user’s input. Upon application startup, the user is prompted to select the standard forms from a collection of forms using an auto-suggesting drop-down menu. Then, the user is taken to the chat window.

4.7.2. Processing Request (Step 2)

Once the user enters the prompt (refer to Figure 11), it is sent to the backend server endpoint as a request with the structure.



```
{
“prompt”: <user input>,
“standard_form”: <standard form name>
}
```

Figure 11. How prompt is configured (source: developed by authors).

4.7.3. Load the Data Using a CSVLoader (Step 3)

Langchain’s CSV Loader module loads a CSV (Comma Separated Values) file into a list of documents. Each CSV file represents a Standard Form, and each record corresponds to a single row inside the CSV file. Each row, a clause in this example, is loaded as a separate document, with the column key being named clause.

4.7.4. Split Document into Chunks Using RecursiveCharacterTextSplitter in the Langchain Library (Step 4)

Langchain’s RecursiveCharacterTextSplitter class separates text into several parts by recursively examining characters. It attempts to divide the text into several characters to find one that works. This function is advantageous since it attempts to maintain the location of all semantically important stuff for as long as feasible. In this approach, the RecursiveCharacterTextSplitter class is used to separate the text of each sentence into 500-character chunks with no overlap between each chunk.

4.7.5. OpenAIEmbeddings (Step 5)

OpenAI Embeddings are numerical representations of concepts turned into number sequences, enabling computers to perceive the connections between such concepts more effectively. OpenAI offers a range of embedding models, each tailored to perform well with certain aspects such as text similarity, text search, and code search. In this instance, embeddings are applied to ensure textual consistency. The values for clause_id, clause_name, and responsibility are converted to OpenAIEmbeddings.

4.7.6. Chroma (Step 6)

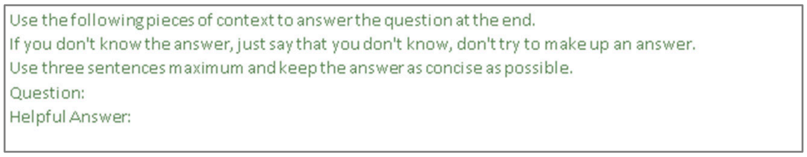
The created embeddings are stored in the chroma.

4.7.7. ChatOpenAI

The principal model is ChatGPT-4.o.

4.7.8. PromptTemplate (Step 7)

A Prompt Template is a class in Langchain that allows you to define the format of the prompts used by a language model when generating text (refer to Figure 12). It provides a way to create dynamic prompts by specifying placeholders for variable content, which will be replaced with actual values when the template is used.



```
Use the following pieces of context to answer the question at the end.
If you don't know the answer, just say that you don't know, don't try to make up an answer.
Use three sentences maximum and keep the answer as concise as possible.
Question:
Helpful Answer:
```

Figure 12. Prompt used (source: developed by authors).

4.7.9. RetrievalQA (Step 8)

RetrievalQA is a module in Langchain that combines a retriever and a QA chain. A retriever is a component that retrieves relevant documents from a collection of documents based on a given query. The retrieved documents are then passed to the QA chain for further processing. A QA chain is a component that takes the retrieved documents and uses them to answer the given query. The QA chain can use various techniques, such as NLP, machine learning, and information retrieval, to generate an answer to the query.

There are several techniques that a retriever can use to retrieve documents, including:

- TF-IDF: This technique matches keywords between the query and the documents, representing them as sparse vectors;
- BM25: This technique matches keywords between the query and the documents;
- Dense embeddings: This technique uses dense embeddings, such as OpenAIEmbeddings, Word2Vec, and BERT, to represent the query and the documents.

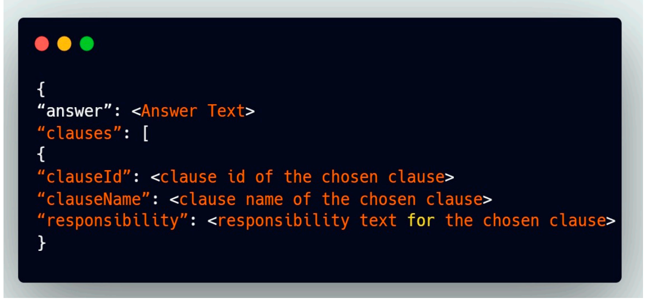
These strategies either employ dense embeddings to represent keywords or compare the keywords in the query with those in the documents. In this case, retrieval was performed via semantic search. All documents are assigned a numerical vector (an embedding) during this procedure, and these vectors are subsequently stored in a vector database (a database optimised for storing and querying vectors).

Vectors are examined and selected based on their resemblance to the query vector in the semantic search. Vectorisation is the process of encoding information about an item into vectors. Vector search operates by encoding information about an object into vectors and then comparing vectors to identify the most similar ones. This function is achieved by converting unstructured data, such as text and images, into a numerical representation using machine learning approaches that capture the meaning and context of unstructured data.

In summary, appropriate clause embeddings are retrieved using semantic search and incorporated into the custom template along with the user prompt during this phase. Then, it is transmitted to the ChatGPT model along with the key and clause ID to obtain a response.

4.7.10. Retrieve Source Clauses to Create the Response (Step 9)

The Pandas library is used to populate a table with data on clauses. The original "clause name" and "responsibility" are obtained using the "clause id", which is unique to each clause. QAChain combines answers with the collected documents to create a JSON object representing the structure (refer to Figure 13). Here, the "clauses" list comprises all the documents utilised. Additionally, the list is sorted by relevance (The clause with the strongest relationship to the preceding clause will be shown first).



```
{
“answer”: <Answer Text>
“clauses”: [
{
“clauseId”: <clause id of the chosen clause>
“clauseName”: <clause name of the chosen clause>
“responsibility”: <responsibility text for the chosen clause>
}
```

Figure 13. Retrieval of JSON object (source: developed by authors).

4.8. Training

While the model’s architecture and underlying algorithms influence the model’s responses, it is important to note that the training data also plays a significant role in determining its behaviour. The initial step in the data collection process involves identifying pertinent sources from which to gather data. In the realm of contract interpretation, precision is paramount. Therefore, sources must be carefully selected. As this is a function practiced by contract managers in construction, the most appropriate source of data is professional data. With the data meticulously collected and rigorously cleaned, the next imperative step is data formatting. Proper formatting is the conduit through which the model effectively learns from the data and produces correct and contextually relevant responses.

Two conventional formats for training conversational AI models are conversational pairs and single input–output sequences. The former comprises paired input messages or prompts with corresponding output responses, suitable for chat-based interactions. The latter stitches together conversational turns into a unified input–output sequence, ideal for scenarios where the model is expected to generate complete dialogues. The conversational pairs method trains the model to produce accurate outputs. In chat-based training, where ChatGPT generates responses based on user inputs, it is imperative to define a clear and structured input–output format. This format governs how data is presented to the model and how it generates responses. System messages, user-specific information, and context preservation must be carefully incorporated to provide unambiguous instructions to the model during training, ensuring that it responds lucidly and contextually appropriately.

5. Discussion

The following is the evaluation of the original ChatGPT response to contractual questions compared to the responses provided by the custom-developed model.

5.1. Evaluation Method

The chatbot model was evaluated using both manual and automated techniques.

Automatic evaluation was used to determine the output text’s quality quickly. Two hundred questions-and-answer pairs were created in order to conduct an automated evaluation. In this study, comparing ChatGPT with a chatbot developed using ChatGPT as a base model, cosine similarity is the most suitable metric for evaluating the performance and semantic alignment between the two models.

Cosine similarity focuses on the directional alignment of embeddings, making it ideal for capturing the semantic similarity of generated text or responses without being affected by the magnitude of the embeddings [56]. This is relevant where the meaning of the text is encoded in the angular relationship between vectors rather than their absolute lengths [57]. By using cosine similarity, the study can effectively compare how closely the outputs of the two models align in terms of meaning and intent, regardless of any variation in vector scale [58]. Studies such as Lahitani, Permanasari and Setiawan [56] demonstrate that cosine similarity excels in these tasks, particularly when applied to word and sentence embeddings. Its robustness and ability to isolate semantic differences make it the most appropriate metric for assessing the semantic consistency and language understanding capabilities between ChatGPT and the custom chatbot.

The model's responses and subsequent responses are converted to OpenAIEmbeddings. Then, the cosine distance is calculated between the model output response and the predefined response. Afterwards, the model is evaluated based on its mean.

$$error = \frac{\Sigma(\cos i\,ne\,dis \tan c\,e)}{100} \tag{2}$$

In the manual procedure, the quality of the produced text is evaluated by human assessors. This can give subtle and thorough feedback on the model's performance but can be time-consuming and costly. This can be overcome by providing the evaluators with a series of questions and requesting that they score the relevance, coherence, and fluency of the generated replies on a specified scale.

In the following study, each final response to the prompts evaluated is graded on a scale from 0 to 4: 0 for irrelevant, 1 for somewhat relevant, 2 for neutral, 3 for highly relevant, and 4 for extremely relevant. Thereinafter, the average score is determined.

$$score = \frac{\Sigma\, score}{100} \tag{3}$$

Each chatbot response has around three sources for each suitable sentence. The clauses are ordered in descending order of relevancy.

$$score = \frac{\Sigma\, score\ of\ 1st\ Source\ +\ Score\ of\ 2nd\ Source\ +\ Score\ of\ 3rd\ Source}{100}$$

Therefore, the manual average score was the average of each prompt out of two hundred prompts marked by three contract interpretation professionals from 1 to 4 based on the relevancy to the topic.

5.2. Distribution of Output Scores

The 200 prompts were extracted from the questions that users were asked, and they were sent to the experts to identify the prominent contractual clauses relevant to the prompt. Those clauses as vector embeddings were compared against the response that the two LLMs generated. Figure 14 illustrates the distribution of cosine similarity values for the 200 responses obtained from the LLM.

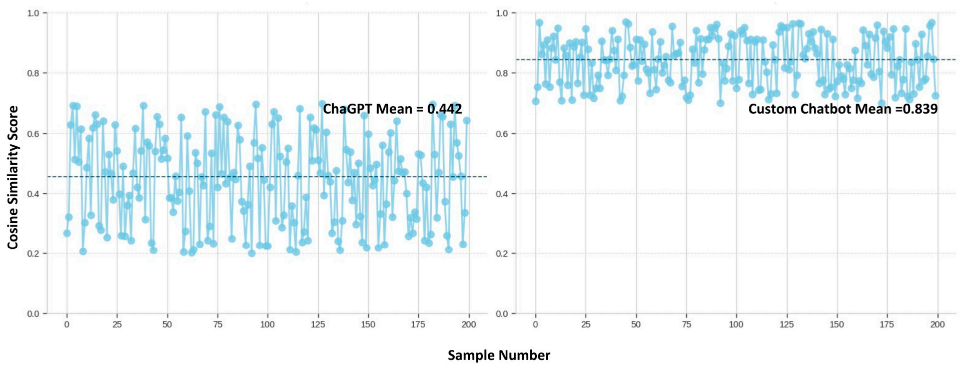


Figure 14. Cosine similarity index scores. The dotted line represents the mean cosine similarity score among 200 responses, and the blue points represent the data points (source: developed by authors).

Then, for the manual evaluation, reach responses generated from the two LLM were sent to three contract interpretation professionals, scoring the responses based on the relevancy of the information provided. The average of these scores was considered for the manual evaluation. Figure 15 shows the distribution of these scores.

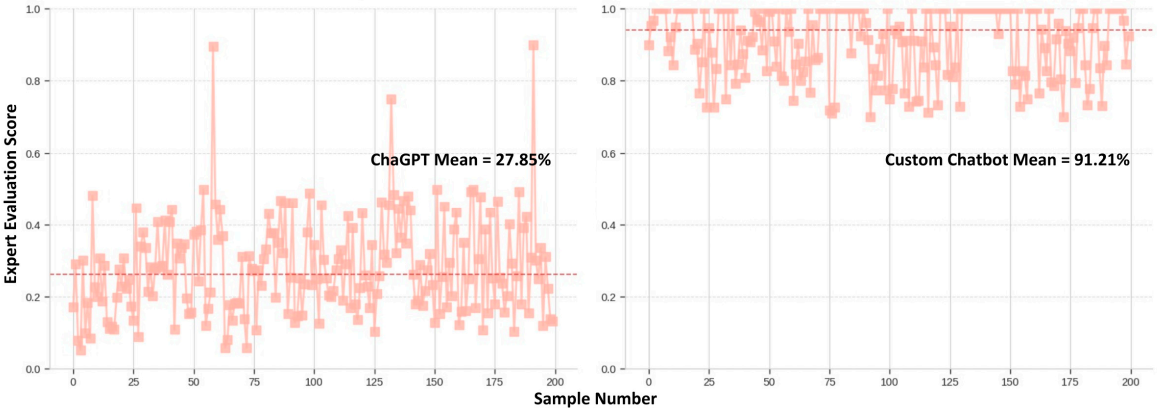


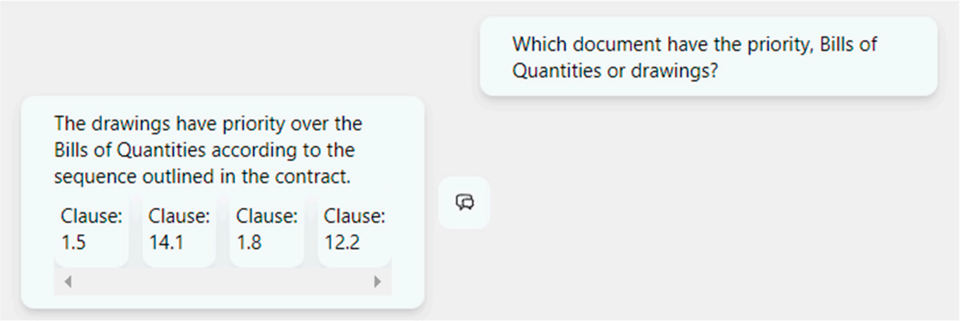
Figure 15. Expert evaluation scores. Dotted line represents the mean cosine similarity score among 200 responses and the red points represent the data points (source: developed by authors).

5.3. Sample Evaluations of the Experts

The following section presents some evaluated example prompts among the two hundred assessed prompts. This will highlight the reasons behind the distribution of the scores and how these models generated outputs for these prompts.

5.3.1. Prompt 1 (Refer to Figure 16): Which Document Prioritises Order, Drawings, or Bills of Quantities?

Figure 16 shows the custom-made chatbot's response to the above prompt.



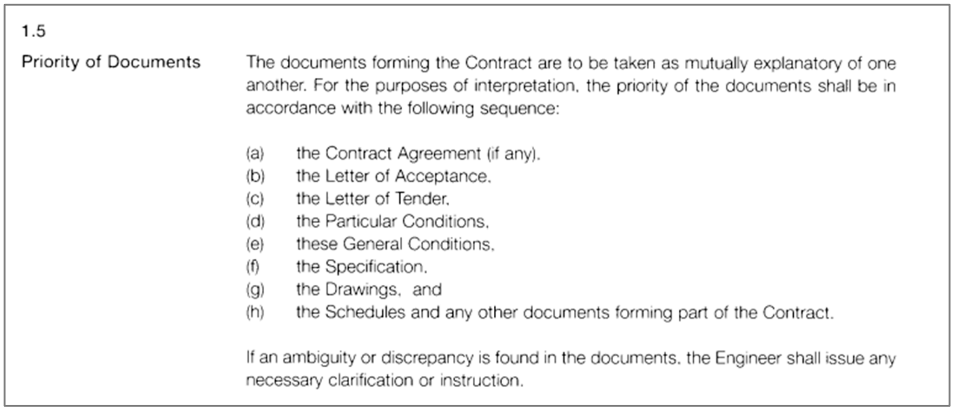
Which document have the priority, Bills of Quantities or drawings?

The drawings have priority over the Bills of Quantities according to the sequence outlined in the contract.

Clause: 1.5 Clause: 14.1 Clause: 1.8 Clause: 12.2

Figure 16. Response to sample Prompt 1 from the chatbot developed (source: developed by authors).

The response refers to Clause 1.5, and the appropriate clause for the FIDIC 1999 is Clause 1.5—Priority of Documents (refer to Figure 17). As mentioned in Figure 17, Clause 1.5 gives a clear answer to the prompt. The other three clauses can be put in order of relevance: 1.8—Care and Supply of Documents, 12.2—Method of Measurement, and 14.1—Contract Price. Thus, the relevance to the prompt theme for the latter two clauses is questionable. Hence, the average score for this sample of Prompt 1 was 75%.



1.5

Priority of Documents

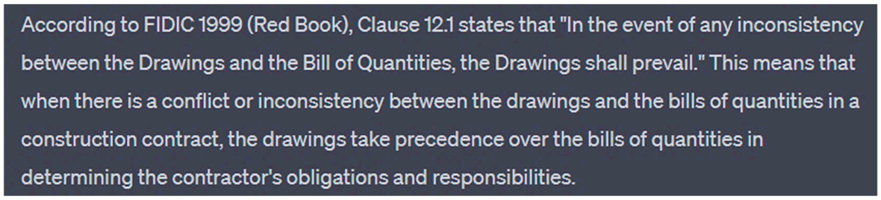
The documents forming the Contract are to be taken as mutually explanatory of one another. For the purposes of interpretation, the priority of the documents shall be in accordance with the following sequence:

(a) the Contract Agreement (if any),
(b) the Letter of Acceptance,
(c) the Letter of Tender,
(d) the Particular Conditions,
(e) these General Conditions,
(f) the Specification,
(g) the Drawings, and
(h) the Schedules and any other documents forming part of the Contract.

If an ambiguity or discrepancy is found in the documents, the Engineer shall issue any necessary clarification or instruction.

Figure 17. Clause 1.5 in FIDIC 1999 [59].

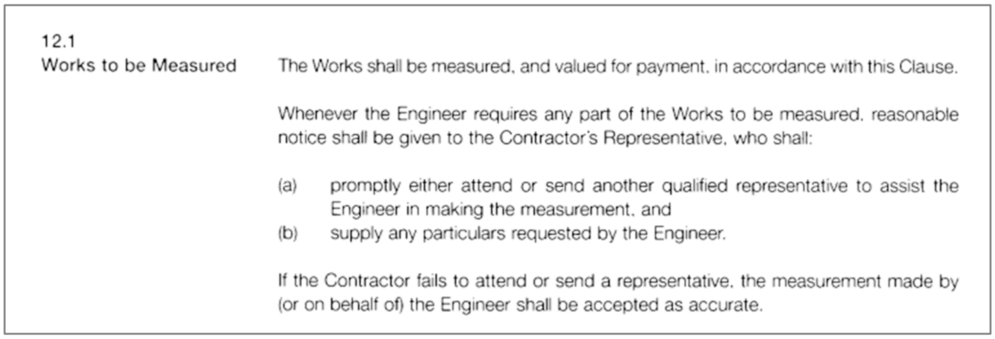
Figure 18 shows the response to the same prompt from the ChatGPT.



According to FIDIC 1999 (Red Book), Clause 12.1 states that "In the event of any inconsistency between the Drawings and the Bill of Quantities, the Drawings shall prevail." This means that when there is a conflict or inconsistency between the drawings and the bills of quantities in a construction contract, the drawings take precedence over the bills of quantities in determining the contractor's obligations and responsibilities.

Figure 18. Response from ChatGPT for sample Prompt 1 (source: ChatGPT 3.5 generated).

The clause referred to in the response as 12.1 is “Works to be Measured”. As shown in Figure 19, the direct quotation “In the event of any inconsistency between the Drawings and Bill of Quantities…” is not a clause in the FIDIC 1999 book. It is a standard error of hallucination in ChatGPT. The average score given to this sample prompt by the ChatGPT response is 32%.



12.1

Works to be Measured

The Works shall be measured, and valued for payment, in accordance with this Clause.

Whenever the Engineer requires any part of the Works to be measured, reasonable notice shall be given to the Contractor's Representative, who shall:

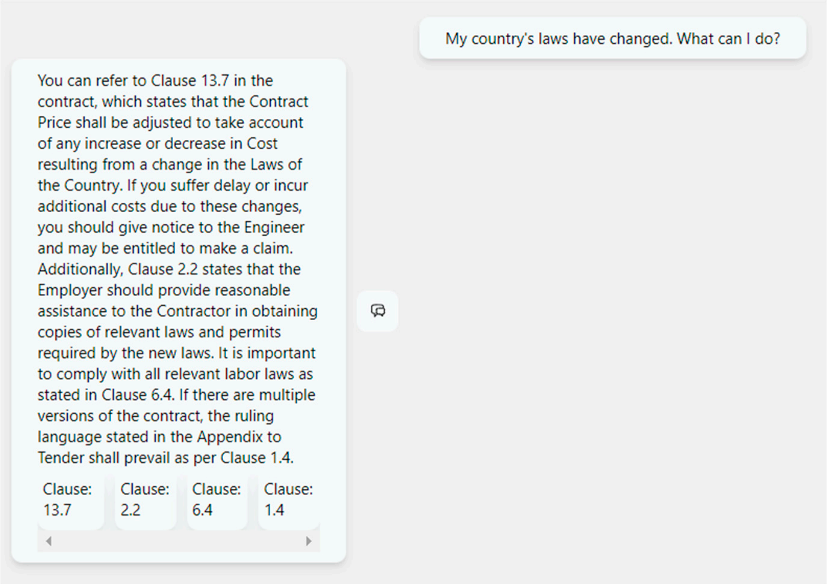
(a) promptly either attend or send another qualified representative to assist the Engineer in making the measurement, and
(b) supply any particulars requested by the Engineer.

If the Contractor fails to attend or send a representative, the measurement made by (or on behalf of) the Engineer shall be accepted as accurate.

Figure 19. Clause 12.1 in FIDIC 1999 [59].

5.3.2. Prompt 2: My Country’s Laws Have Changed: What Can I Do?

Figures 20 and 21 compare two replies to the query. The custom chatbot delivers an exhaustive response to the question with the relevant clauses. The ChatGPT’s response is wrong, as there are clauses specifically related to the country’s laws. Thus, the chatbot’s average score is 85%, while the ChatGPT’s answer is 0%.

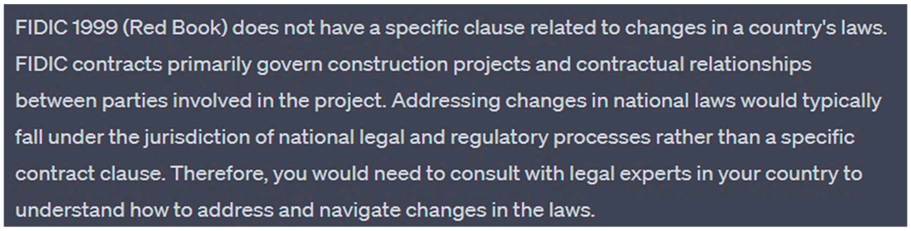


My country's laws have changed. What can I do?

You can refer to Clause 13.7 in the contract, which states that the Contract Price shall be adjusted to take account of any increase or decrease in Cost resulting from a change in the Laws of the Country. If you suffer delay or incur additional costs due to these changes, you should give notice to the Engineer and may be entitled to make a claim. Additionally, Clause 2.2 states that the Employer should provide reasonable assistance to the Contractor in obtaining copies of relevant laws and permits required by the new laws. It is important to comply with all relevant labor laws as stated in Clause 6.4. If there are multiple versions of the contract, the ruling language stated in the Appendix to Tender shall prevail as per Clause 1.4.

Clause: 13.7 Clause: 2.2 Clause: 6.4 Clause: 1.4

Figure 20. Response to sample Prompt 2 from the chatbot developed.

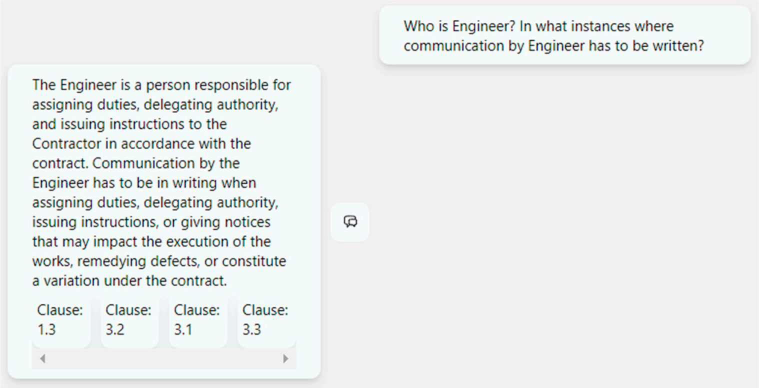


FIDIC 1999 (Red Book) does not have a specific clause related to changes in a country's laws. FIDIC contracts primarily govern construction projects and contractual relationships between parties involved in the project. Addressing changes in national laws would typically fall under the jurisdiction of national legal and regulatory processes rather than a specific contract clause. Therefore, you would need to consult with legal experts in your country to understand how to address and navigate changes in the laws.

Figure 21. Response from ChatGPT for sample Prompt 2.

5.3.3. Prompt 3: Who Is an Engineer? In What Instances Does Communication by the Engineer Have to Be Written?

The responses given by ChatGPT and the chatbot are presented in the two above figures. ChatGPT's answer to this prompt is substantially accurate. The chatbot delivers an accurate response to the question, although it could be comprehensive. Therefore, the average score of the chatbot is 68%, and the score of ChatGPT is 65%. Although ChatGPT's answer does not contain all relevant clauses, it provides an equally comprehensive answer. Therefore, as the score determines the distance between the actual answer and the answer provided by the model, it can be concluded that both answers are at an equal distance (Figures 22 and 23).

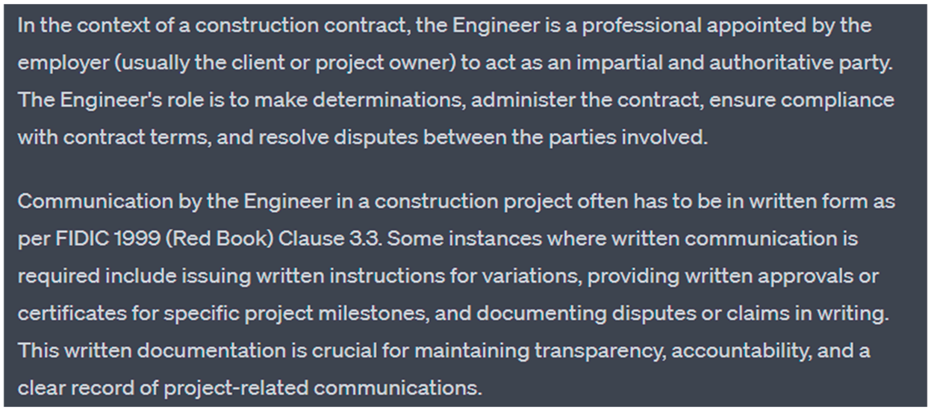


Who is Engineer? In what instances where communication by Engineer has to be written?

The Engineer is a person responsible for assigning duties, delegating authority, and issuing instructions to the Contractor in accordance with the contract. Communication by the Engineer has to be in writing when assigning duties, delegating authority, issuing instructions, or giving notices that may impact the execution of the works, remedying defects, or constitute a variation under the contract.

Clause: 1.3 Clause: 3.2 Clause: 3.1 Clause: 3.3

Figure 22. Response to sample Prompt 3 from the Chatbot developed. (source: screen capture of the chatbot developed)



In the context of a construction contract, the Engineer is a professional appointed by the employer (usually the client or project owner) to act as an impartial and authoritative party. The Engineer's role is to make determinations, administer the contract, ensure compliance with contract terms, and resolve disputes between the parties involved.

Communication by the Engineer in a construction project often has to be in written form as per FIDIC 1999 (Red Book) Clause 3.3. Some instances where written communication is required include issuing written instructions for variations, providing written approvals or certificates for specific project milestones, and documenting disputes or claims in writing. This written documentation is crucial for maintaining transparency, accountability, and a clear record of project-related communications.

Figure 23. Response from ChatGPT for sample Prompt 3 (source: ChatGPT generated).

5.4. Conclusive Evaluation

The ChatGPT base model achieved a cosine similarity score of 0.442 and a 27.85% accuracy in the expert scoring methods. It could be concluded that the average score was 36%. The following violin plot shows the final scores of the ChatGPT base model. As shown in Figure 24, the automatic evaluation scores lie within a closed range compared to the manually evaluated scores.

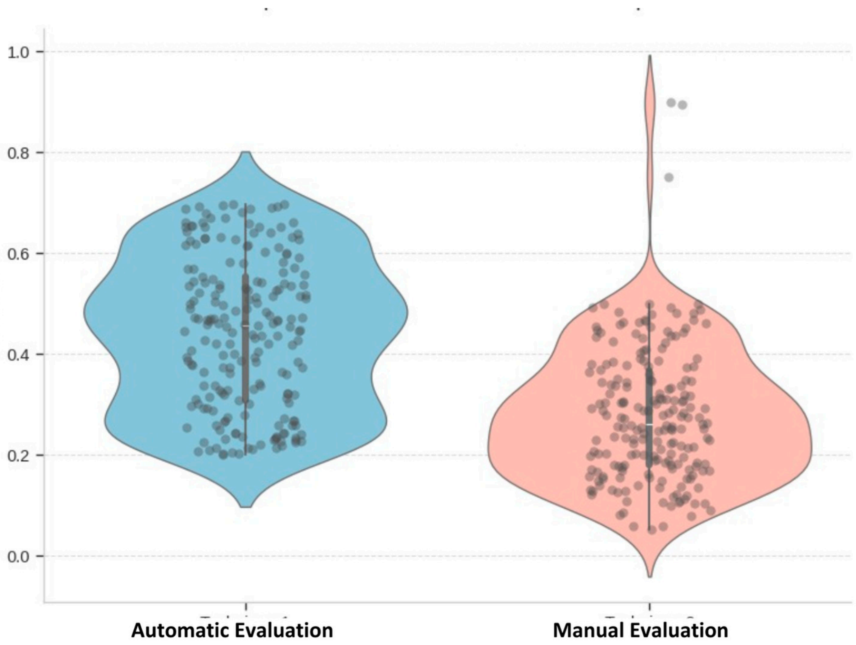


Figure 24. Violin plot of evaluated scores of the ChatGPT base model (source: developed by authors).

The specialised custom chatbot scored 0.839 in the automatic evaluation and 91.21%. Therefore, the average score for the model is 87.56%. The violin plot, shown in Figure 25, is the distribution of these scores for the custom-made chatbot's responses.

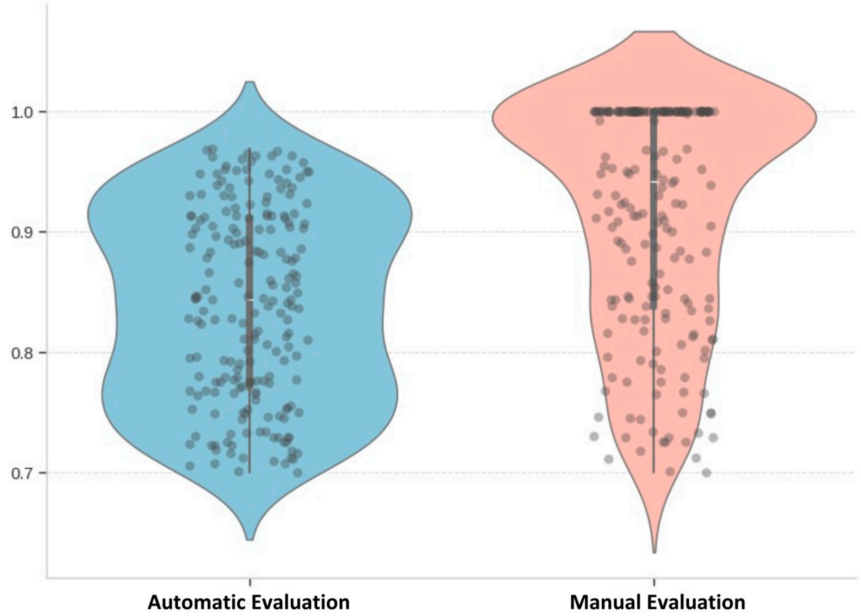


Figure 25. Violin plot of evaluated scores of the custom-made chatbot model (source: developed by authors).

The expert evaluation scores showed significant variation, ranging from 100% accuracy for most prompts generated by the custom chatbot to as low as 10% for those produced by the ChatGPT model. This suggests the custom chatbot's responses were highly relevant and met the experts' expectations. Furthermore, this could imply that ChatGPT's responses were not as appropriate or effective in capturing the experts' attention. However, this might be motivated by the immediate clause specification for the custom chatbot's answer, which immediately gains the reader's attention. In addition, human experts appeared to favour extreme scores, choosing the highest or lowest options for relevancy more frequently. This behaviour could reflect their subjective preferences and biases when evaluating the responses. The experts' inclination towards extreme scores might indicate that they were particularly impressed or unimpressed with certain responses, leading them to rate those responses as either excellent or poor.

The automatic evaluation scores are distributed in a closer range and more distributed to all values. The automatic evaluation scores were more evenly distributed across a range of values. This suggests the automated evaluation methods were less extreme in their assessments and provided a more balanced view of the chatbots' performance. One potential drawback of automatic evaluations is that they may not fully capture actual users' nuanced preferences and subjective judgments. While automatic methods can provide consistent and unbiased assessments, they might overlook specific aspects that human users find important or valuable. Therefore, an average of these final scores must be assessed for a comprehensive, unbiased, and context-aware evaluation.

In conclusion, the specialised chatbot outperformed ChatGPT in the task of construction contract interpretation, with an average score of 88% across 200 responses, compared to ChatGPT's 36%. This significant difference highlights the specialised chatbot's superior ability to provide accurate and context-specific interpretations of contract terms. In contrast, ChatGPT's broader, general-purpose model may lack the domain-specific knowledge required for precise legal interpretation. The results indicate that a chatbot trained specifically for construction contract analysis is significantly better equipped to handle the complexities of a specialised task. The specialised chatbot's higher performance likely stems from its targeted training in construction law and contract terminology, enabling it to understand and apply nuanced legal principles more effectively. This highlights the importance of domain-specific AI tools in contract interpretation, where precision and expertise are critical to achieving reliable outcomes.

Chalkidis et al. [60] highlighted the importance of tailoring LLMs to specific legal domains using the infamous Legal-BERT model. This study will adapt this model to the legal construction domain. While some studies, as mentioned in the literature, focus on broader construction management challenges, this study addresses the issue of contractual ambiguities related to legal clauses outlined in legal guidebooks. Furthermore, incorporating domain-specific training into LLMs enhances performance and sets a new benchmark for specialised legal AI applications.

5.5. Output Validation

Overall, from the total of 200 responses, the average score of the specialised chatbot was 88%, while the average score of ChatGPT was 36%. As this score was obtained by calculating the cosine similarity, a recognised similarity measurement was used to evaluate LLMs; therefore, the accuracy of the value can be ascertained. Furthermore, the cosine distances were calculated from a chatbot developed; thus, the data derived were actual data rather than based on perceptions.

In addition, the output was evaluated by three human experts, and an average was calculated. In addition to minimising the limitations of each evaluation method, an average of these scores was used to obtain a holistic evaluation measure.

5.6. Novelty and Contributions of the Study

The findings of this study highlight the significant advantages of employing domain-specific LLMs in contract interpretation within the construction industry. The comparison between the custom-built chatbot and the baseline ChatGPT model underscores the importance of adapting AI-driven solutions to industry-specific requirements. The specialised chatbot achieved substantially higher accuracy than ChatGPT, demonstrating its superior performance in correctly interpreting construction contract clauses. The development project presented in this paper aligns with previous research that emphasised the necessity of tailoring LLMs to specific domains for enhanced accuracy and reliability [60].

This study presents a novel solution for developing a customised model with vector databases containing contract clauses. This is a timely approach when AI agents are the core of industry interest at the time of this publication. This paper also focuses on the construction field, where previous studies have focused primarily on automating the construction contracts, or applying LLMs in the construction industry without refining them for contract interpretation. For example, Saka et al. [61] and Prieto, Mengiste, and García de Soto [7] explored LLMs for document generation, including contract agreements and claim documentation, but did not extensively tailor them to contract-specific queries.

This study contributes to knowledge in several key ways. Firstly, it suggests user requirements and the obstacles inherent in contract question-answering using an LLM. Secondly, it proposes a workflow for designing chatbots capable of effectively addressing contractual queries. By doing so, the research in the field of technology adoption advances the use of intelligent contracts. This is specifically important in complex projects where many legal issues arise due to the involvement of many international stakeholders, so guiding project managers is critical to reducing disputes. Finally, it delivers a tangible proof-of-concept chatbot specialised in contract-related responses, furthering practical AI applications in the construction industry.

5.7. Performance Evaluation and Validation

The cosine similarity performance evaluation further validates the efficacy of embedding-based retrieval methods in legal contract analysis. Prior work in AI for law [52] has shown that vector databases significantly improve the retrieval of relevant legal documents. Our findings confirm this trend, illustrating how embeddings facilitate context-aware contract interpretation, bridging the gap between AI language models and domain-specific applications. This advancement supports the argument that future contract analysis systems should integrate AI with domain-specific knowledge bases to improve precision.

Despite the promising results, this study also highlights the limitations of current AI-based contract interpretation methods. Human oversight remains essential, as evidenced by variations in expert evaluations, where human assessors exhibited biases towards extreme scoring. This aligns with findings from prior studies [56] that suggest human evaluators often perceive AI-generated text either as highly accurate or entirely incorrect, depending on their expectations and familiarity with the technology. Future research should explore hybrid approaches that incorporate both AI-driven recommendations and human verification to ensure robust contract analysis.

5.8. Further Research Directions

The integration of big data analytics and blockchain, with optimal LLMs, has the potential to enhance the efficiency of contract managements and interpretations in the construction industry. Big data provides historical insights for risk assessment, while blockchain ensures transparency in contractual terms, reducing disputes. Future studies could also look into the use of next-generation AI agents, especially those with smart planning abilities, like reinforcement learning or hierarchical task networks. These AI tools have the potential to review construction contracts for any non-compliance issues and help teams keep a closer eye on financial commitments and project costs. The AI agents should be able to conduct risk assessment by identifying missing clauses, analysing payment terms of the construction project, comparing against benchmarks in the previous projects, and executing them autonomously.

As another future direction, expanding LLMs to include computer vision and the ability to interpret technical drawings or the quality of the installed material at the construction site can enhance project management and reduce errors. AI agents can further streamline processes like preventing dispute resolution during the project by swiftly analysing contract clauses, national regulations, company rules, and offering data-driven insights to project managers. This helps real-time compliance checks systems to be designed and adjusted to each project.

These systems or AI agents should also be evaluated in terms of reliability and legal and ethical implications, and the outcome of the initial versions of these systems can be cross-verified by senior experts. Developing AI agents capable of analysing contracts in multiple languages would enhance global applicability, especially in international tunnelling, tower, dam, or bridge projects with various parties from multiple countries. The literature supports the value of intelligent contracts in the construction industry, which suffers from significant contractual disputes. Integrating chatbots with BIM and GIS has the potential to evolve current conversational systems into more intelligent, legally accurate, and autonomous tools. These advanced AI systems could understand and interpret information more perceptively while also factoring in local laws, construction standards, stakeholder responsibilities, and internal organisational policies.

5.9. Recommendations for Industry Adoption

While models such as ChatGPT demonstrate substantial accuracy, the complexity of construction contract language necessitates further refinement for optimal interpretation.

The outputs can also be validated or triangulated using other tools. Future research should focus on the scientific evaluation of these bots and systems and also explore integrating construction-specific ontologies with vector databases. This helps to create a more robust framework for automating complex document analysis, risk assessment, and decision-making in construction.

Additionally, development efforts should focus on evaluating the ethical and practical implications of increased automation within the industry. Ensuring that these tools augment human expertise rather than replace it is crucial. Collaborative efforts between academia and industry practitioners on case studies or using action research methods would provide valuable insights and further refine these models to align with real-world requirements.

Construction companies should actively seek solutions that integrate LLMs like ChatGPT into their operations to streamline reporting, contract interpretation, and risk management processes. Implementing models trained on industry-specific data can result in substantial time-saving and improved decision-making accuracy. To maximise the benefits of AI-powered technologies, firms should invest in developing models that enable more efficient and intelligent document retrieval systems, allowing stakeholders to quickly access relevant insights and proactively manage risks.

However, human oversight must remain a priority to address inaccuracies and ensure output quality. Regular audits of AI-assisted decisions will help maintain trust in these systems and optimise their application in dynamic construction environments. Furthermore, industry-wide adoption will require regulatory and ethical considerations. Issues such as liability for AI-generated misinterpretations, data privacy concerns, and regulatory acceptance of AI-driven contract analysis must be addressed. Collaboration between legal professionals, AI researchers, and policymakers is necessary to establish standards and guidelines for the responsible deployment of AI in contract interpretation.

6. Conclusions

The implications of employing LLMs within the construction sector are manifold. This paper presents a model that can be useful in contract management. Through proper modifications, the model can adeptly navigate the nuances of construction-related language, automating reports and contract interpretation with potential time-savings of up to 70% and an accuracy of 85%. Although the model needs further tests on various contract types and datasets, this efficiency shows that the practice is useful to stakeholders with timely insights, fostering informed decision-making. Moreover, LLMs and AI agents focusing on contract management will have advanced analytical abilities to enhance risk assessment and empower project managers with proactive risk mitigation. They excel in tasks such as information extraction, content summarisation, and answering document-related queries. For instance, a chatbot can handle questions like, "What happens if the agreement is not signed?" or "Identify risks associated with this project".

A mechanism integrating vector databases and embeddings is essential to enable these functions. Vector databases manage vector data, where embeddings represent words or phrases as points in a multi-dimensional space. This allows language models to understand meaning based on relative positions, enabling efficient data search and comparison.

Unlike traditional keyword searches that rely on exact matches, language models use vector similarity to find documents with semantically similar meanings, even if specific words do not match the query. This revolutionises the search for unstructured data, revealing insights that conventional methods might miss.

ChatGPT and chatbot responses demonstrate this technology's potential, as shown in the previous development section. ChatGPT scored 36% accuracy, while the chatbot scored 88%. Across 200 responses, the difference between the two models' accuracies highlights both the effectiveness and areas for improvement in applying language models to construction. Therefore, it can be concluded that developing a custom-built chatbot with ChatGPT is the optimal solution for utilising large LLMs and ChatGPT in the construction industry.

Author Contributions: Conceptualisation, P.V.I.N.S. and S.S.; methodology: P.V.I.N.S. and S.S.; software development: P.V.I.N.S.; writing—original draft: all authors; data curation: P.V.I.N.S.; visualisations: P.V.I.N.S. and S.M.E.; specific knowledge: S.S.; writing—review and editing, all authors; aiding software development and funding acquisition, H.S.J. and B.A.I.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the World Bank AHEAD project fund received through the University Business Linkage Cell, University of Moratuwa. The funding reference is UOM/UBLC/MAG/CER/01.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Amoah, C.; Nkosazana, H. Effective management strategies for construction contract disputes. *Int. J. Build. Pathol. Adapt.* **2023**, *41*, 70–84. https://doi.org/10.1108/IJBPA-01-2022-0004.
2. Sigalov, K.; Ye, X.; König, M.; Hagedorn, P.; Blum, F.; Severin, B.; Hettmer, M.; Hückinghaus, P.; Wölkerling, J.; Groß, D. Automated Payment and Contract Management in the Construction Industry by Integrating Building Information Modeling and Blockchain-Based Smart Contracts. *Appl. Sci.* **2021**, *11*, 7653.
3. KPMG. *Global Construction Survey 2019: Future-Ready Index*; KPMG: New York, NY, USA, 2019.
4. Kumar Viswanathan, S.; Panwar, A.; Kar, S.; Lavingiya, R.; Jha, K.N. Causal modeling of disputes in construction projects. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2020**, *12*, 04520035.
5. Phongwattana, T.; Chan, J.H. Automated Extraction and Visualization of Metabolic Networks from Biomedical Literature Using a Large Language Model. *bioRxiv* **2023**. https://doi.org/10.1101/2023.06.27.546560.
6. Locatelli, M.; Seghezzi, E.; Pellegrini, L.; Tagliabue, L.C.; Di Giuda, G.M. Exploring natural language processing in construction and integration with building information modeling: A scientometric analysis. *Buildings* **2021**, *11*, 583.
7. Prieto, S.A.; Mengiste, E.T.; García de Soto, B. Investigating the Use of ChatGPT for the Scheduling of Construction Projects. *Buildings* **2023**, *13*, 857.
8. Aluga, M. Application of CHATGPT in civil engineering. *East Afr. J. Eng.* **2023**, *6*, 104–112.
9. Morgan, D.L. Pragmatism as a paradigm for social research. *Qual. Inq.* **2014**, *20*, 1045–1053.
10. Uher, T.E.; Uher, T.; Davenport, P. *Fundamentals of Building Contract Management*; UNSW Press: Randwick, Australia, 2009.
11. Aladağ, H. Assessing the accuracy of ChatGPT use for risk management in construction projects. *Sustainability* **2023**, *15*, 16071.
12. Rameezdeen, R.; Rodrigo, A. Modifications to standard forms of contract: The impact on readability. *Australas. J. Constr. Econ. Build.* **2014**, *14*, 31–40.
13. Kreye, M.; Balangalibun, S. Uncertainty in project phases: A framework for organisational change management. In Proceedings of the 15th Annual Conference on the European Academy of Management, Warsaw, Poland, 17–20 June 2015.
14. Andrews, N. Interpretation of contracts and "commercial common sense": Do not overplay this useful criterion. *Camb. Law J.* **2017**, *76*, 36–62.
15. Global Construction Perspectives. *Global Construction 2030: A Global Forecast for the Construction Industry to 2030*; Global Construction Perspectives: London, UK, 2023.
16. Zhang, X.; Yang, S.; Duan, L.; Lang, Z.; Shi, Z.; Sun, L. Transformer-XL with graph neural network for source code summarisation. In Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 17–20 October 2021; pp. 3436–3441.
17. Ding, Y.; Ma, J.; Luo, X. Applications of natural language processing in construction. *Autom. Constr.* **2022**, *136*, 104169. https://doi.org/10.1016/j.autcon.2022.104169.

18. Wu, C.; Li, X.; Guo, Y.; Wang, J.; Ren, Z.; Wang, M.; Yang, Z. Natural language processing for smart construction: Current status and future directions. *Autom. Constr.* **2022**, *134*, 104059. https://doi.org/10.1016/j.autcon.2021.104059.
19. Jin, Y.; Cheng, K.; Wang, X.; Cai, L. A Review of Text Sentiment Analysis Methods and Applications. *Front. Bus. Econ. Manag.* **2023**, *10*, 58–64. https://doi.org/10.54097/fbem.v10i1.10171.
20. Chen, Y.; Liu, B.; Wang, X. Automatic text summarisation based on textual cohesion. *J. Electron.* **2007**, *24*, 338–346.
21. Shamshiri, A.; Ryu, K.R.; Park, J.Y. Text mining and natural language processing in construction. *Autom. Constr.* **2024**, *158*, 105200.
22. Zou, Y.; Kiviniemi, A.; Jones, S.W. Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Autom. Constr.* **2017**, *80*, 66–76.
23. Shekhar, G.; Bodkhe, S.; Fernandes, K. On-Demand Intelligent Resource Assessment and Allocation System Using NLP for Project Management. *AMCIS 2020 Proceedings*, 2020. Available online: https://aisel.aisnet.org/amcis2020/it_project_mgmt/it_project_mgmt/8 (accessed on 29 December 2024).
24. Jiang, H.; Lin, P.; Qiang, M. Public-opinion sentiment analysis for large hydro projects. *J. Constr. Eng. Manag.* **2016**, *142*, 05015013.
25. Xue, X.; Hou, Y.; Zhang, J. Automated construction contract summarisation using natural language processing and deep learning. In Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC), Bogota, Colombia, 12–15 July 2022; pp. 459–466.
26. Xu, N.; Zhou, X.; Guo, C.; Xiao, B.; Wei, F.; Hu, Y. Text Mining Applications in the Construction Industry: Current Status, Research Gaps, and Prospects. *Sustainability* **2022**, *14*, 16846.
27. Rane, N.; Choudhary, S.; Rane, J. Integrating ChatGPT, Bard, and Leading-Edge Generative Artificial Intelligence in Building and Construction Industry: Applications, Framework, Challenges, and Future Scope. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4645597 (accessed on 29 December 2024).
28. Li, R.Y.M.; Li, R.Y.M. Automated and intelligent tools in the construction industry. In *Construction Safety Informatics*; Springer: Singapore, 2019; pp. 103–119.
29. He, C.; Yu, B.; Liu, M.; Guo, L.; Tian, L.; Huang, J. Utilising Large Language Models to Illustrate Constraints for Construction Planning. *Buildings* **2024**, *14*, 2511.
30. Pavlik, J.V. Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *J. Mass Commun. Educ.* **2023**, *78*, 84–93.
31. Liu, Y.; Han, T.; Ma, S.; Zhang, J.; Yang, Y.; Tian, J.; He, H.; Li, A.; He, M.; Liu, Z. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-radiology* **2023**, *1*, 100017.
32. Koroteev, M.V. BERT: A review of applications in natural language processing and understanding. *arXiv* **2021**, arXiv:2103.11943.
33. Shreyashree, S.; Sunagar, P.; Rajarajeswari, S.; Kanavalli, A. A literature review on bidirectional encoder representations from transformers. In *Inventive Computation and Information Technologies: Proceedings of ICICIT 2021*; Springer: Singapore, 2022; pp. 305–320.
34. Rajapaksha, P.; Farahbakhsh, R.; Crespi, N. Bert, xlnet or roberta: The best transfer learning model to detect clickbaits. *IEEE Access* **2021**, *9*, 154704–154716.
35. Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.; Le, Q.V.; Salakhutdinov, R. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv* **2019**, arXiv:1901.02860.
36. Yuan, S.; Zhao, H.; Du, Z.; Ding, M.; Liu, X.; Cen, Y.; Zou, X.; Yang, Z.; Tang, J. Wudaocorpora: A super large-scale chinese corpora for pre-training language models. *AI Open* **2021**, *2*, 65–68.
37. Raiaan, M.A.K.; Mukta, M.S.H.; Fatema, K.; Fahad, N.M.; Sakib, S.; Mim, M.M.J.; Ahmad, J.; Ali, M.E.; Azam, S. A review on large Language Models: Architectures, applications, taxonomies, open issues and challenges. *IEEE Access* **2024**, *12*, 26839–26874.
38. Ono, K.; Morita, A. Evaluating large language models: Chatgpt-4, mistral 8x7b, and google gemini benchmarked against mmlu. *Authorea Prepr.* **2024**. https://doi.org/10.36227/techrxiv.170956672.21573677/v1.
39. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y. A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–45.
40. Wang, N.; Issa, R.R.; Anumba, C.J. NLP-based query-answering system for information extraction from building information models. *J. Comput. Civ. Eng.* **2022**, *36*, 04022004.
41. Lee, J.; Yi, J.-S.; Son, J. Development of automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP. *J. Comput. Civ. Eng.* **2019**, *33*, 04019003.
42. Chung, S.; Moon, S.; Kim, J.; Kim, J.; Lim, S.; Chi, S. Comparing natural language processing (NLP) applications in construction and computer science using preferred reporting items for systematic reviews (PRISMA). *Autom. Constr.* **2023**, *154*, 105020.

43. Moon, S.; Lee, G.; Chi, S. Automated system for construction specification review using natural language processing. *Adv. Eng. Inform.* **2022**, *51*, 101495.
44. Rane, N. Role of ChatGPT and Similar Generative Artificial Intelligence (AI) in Construction Industry. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4598258 (accessed on 31 December 2024).
45. Chen, J.-H.; Su, M.-C.; Azzizi, V.T.; Wang, T.-K.; Lin, W.-J. Smart project management: Interactive platform using natural language processing technology. *Appl. Sci.* **2021**, *11*, 1597.
46. Jafari, P.; Al Hattab, M.; Mohamed, E.; AbouRizk, S. Automated extraction and time-cost prediction of contractual reporting requirements in construction using natural language processing and simulation. *Appl. Sci.* **2021**, *11*, 6188.
47. Samsami, R. Optimizing the Utilization of Generative Artificial Intelligence (AI) in the AEC Industry: ChatGPT Prompt Engineering and Design. *CivilEng* **2024**, *5*, 971–1010.
48. Hassan, F.U.; Le, T.; Lv, X. Addressing legal and contractual matters in construction using natural language processing: A critical review. *J. Constr. Eng. Manag.* **2021**, *147*, 03121004.
49. Moshood, T.D.; Adeleke, A.; Nawanir, G.; Mahmud, F. Ranking of human factors affecting contractors' risk attitudes in the Malaysian construction industry. *Soc. Sci. Humanit. Open* **2020**, *2*, 100064.
50. Hassan, F.U.; Le, T.; Le, C. Automated approach for digitalising scope of work requirements to support contract management. *J. Constr. Eng. Manag.* **2023**, *149*, 04023005.
51. Srivastawa, A.K. Exploring Contract Management in the Digital Age: The Impact of Artificial Intelligence. *Jus Corpus LJ* **2023**, *4*, 737.
52. Cappuzzo, R.; Papotti, P.; Thirumuruganathan, S. Creating embeddings of heterogeneous relational datasets for data integration tasks. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, Portland, OR, USA, 14–19 June 2020; pp. 1335–1349.
53. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need.(Nips), 2017. *arXiv* **2017**, arXiv:1706.03762.
54. Topsakal, O.; Akinci, T.C. Creating large language model applications utilising langchain: A primer on developing llm apps fast. In Proceedings of the International Conference on Applied Engineering and Natural Sciences, Konya, Turkey, 10–12 July 2023; pp. 1050–1056.
55. Maas, A.; Daly, R.E.; Pham, P.T.; Huang, D.; Ng, A.Y.; Potts, C. Learning word vectors for sentiment analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 142–150.
56. Lahitani, A.R.; Permanasari, A.E.; Setiawan, N.A. Cosine similarity to determine similarity measure: Study case in online essay assessment. In Proceedings of the 2016 4th International Conference on Cyber and IT Service Management, Bandung, Indonesia, 26–27 April 2016; pp. 1–6.
57. Sohangir, S.; Wang, D. Improved sqrt-cosine similarity measurement. *J. Big Data* **2017**, *4*, 1–13.
58. Qian, G.; Sural, S.; Gu, Y.; Pramanik, S. Similarity between Euclidean and cosine angle distance for nearest neighbor queries. In Proceedings of the 2004 ACM Symposium on Applied Computing, Nicosia, Cyprus, 14–17 March 2004; pp. 1232–1237.
59. International Federation of Consulting Engineers (FIDIC). *Conditions of Contract for Construction*, 1st ed.; International Federation of Consulting Engineers: Geneva, Switzerland, 1999.
60. Chalkidis, I.; Fergadiotis, M.; Malakasiotis, P.; Aletras, N.; Androutsopoulos, I. Legal-Bert: The muppets straight out of law school. *arXiv* **2020**, arXiv:2010.02559.
61. Saka, A.; Taiwo, R.; Saka, N.; Salami, B.A.; Ajayi, S.; Akande, K.; Kazemi, H. GPT models in construction industry: Opportunities, limitations, and a use case validation. *Dev. Built Environ.* **2023**, *17*, 100300.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.