applied sciences

MDPI

Article

In-House Knowledge Management Using a Large Language Model: Focusing on Technical Specification Documents Review

Jooyeup Lee [1], Wooyong Jung [1,*] and Seungwon Baek [2,*]

[1] Department of Nuclear Power Plant Engineering, KEPCO International Nuclear Graduate School, 658-91 Haemaji-ro, Seosaeng-myeon, Ulju-gun, Ulsan 45014, Republic of Korea; kosdac@gmail.com
[2] Post-Construction Evaluation and Management Center, Department of Construction Policy Research, Korea Institute of Civil Engineering and Building Technology, Goyang 10223, Republic of Korea
* Correspondence: trustjung@gmail.com (W.J.); baeksw@kict.re.kr (S.B.)

Abstract: In complex construction projects, technical specifications have to be reviewed in a short period of time. Even experienced engineers find it difficult to review every detail of technical specifications. In addition, it is not easy to transfer experienced knowledge to junior engineers. With the technological innovation of large language models such as ChatGPT, a fine-tuned language model is proposed as an effective solution for the automatic review of technical specification documents. Against this backdrop, this study examines the in-house technical specification documents that are not publicly available. Then, two fine-tuned large language models, GPT-3 and LLaMA2, are trained to answer questions related to technical specification documents. The results show that the fine-tuned LLaMA2 model generally outperforms the fine-tuned GPT-3 model in terms of accuracy, reliability, and conciseness of responses. In particular, the fine-tuned LLaMA2 model suppressed hallucinogenic effects better than the fine-tuned GPT-3 model. Based on the results, this study discussed the applicability and limitations of a fine-tuned large language model for in-house knowledge management. The results of this study are expected to assist practitioners in developing a domain-specific knowledge management solution by fine-tuning an open-source large language model with private datasets.

Keywords: technical specification document; knowledge management; large language model; fine-tuning; GPT; LLaMA

check for updates

Citation: Lee, J.; Jung, W.; Baek, S. In-House Knowledge Management Using a Large Language Model: Focusing on Technical Specification Documents Review. *Appl. Sci.* **2024**, *14*, 2096. https://doi.org/10.3390/app14052096

Academic Editor: Gianluigi Ferrari

Received: 27 January 2024
Revised: 20 February 2024
Accepted: 29 February 2024
Published: 2 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Technological advancements and multinational corporations dominate the global economy in the information age and knowledge society. The complexity and scale of corporations and their technical requirements are driving the need for more effective knowledge management (KM) solutions. Particularly, in complex engineering projects such as oil and gas and nuclear power plants, engineering data and documentation are tremendous and should be reviewed in a short bid period. For example, within two months, the contractors in Qatar construction project should have reviewed the engineering specification, which included more than 4000 pages of requirements and 13,000 references [1]. A comprehensive understanding of such a vast body of knowledge requires significant time, and the interconnected nature of multi-disciplinary teams does not lend itself to intuitive or easy understanding. If a contractor fails to comply with the provisions of the technical specifications, the project is exposed to latent risks such as rework, delays, defects, and accidents [2].

According to a McKinsey study, employees spend nearly 20% of their working week searching for internal information or asking colleagues for help [3]. Knowledge transfer from senior engineers to junior engineers is vital for enhancing an organization's competitiveness. However, the construction industry has suffered from the aging of the workforce including managerial employees [4,5]. The departure of senior professional can lead to significant knowledge gaps, thereby increasing operational risks and posing potential setbacks [6]. This situation is further aggravated by the transitory nature of the project-based industry where temporary project teams are organized, work together for a while, and dissolve after finishing the project [7]. Despite various efforts, due to internal personnel and external environment changes, essential organizational knowledge can fade or even vanish.

To address KM challenges, diverse methodologies, and technologies have been deployed over the years including the strategic capturing, storing, using, and sharing of information throughout the company [8]. Traditional databases, manual documentation, and even computer-based solutions have been beneficial; however, they also have limitations such as time-consuming data retrieval processes [9,10]. With the advancement of natural language processing (NLP) technology, some researchers have attempted to automate the specification document review using NLP [11–14]. Nevertheless, existing NLP-driven specification document review techniques were limited in terms of the inability to provide contextual insights. In particular, traditional approaches typically require the manual creation of a rule set for the model development and have difficulty understanding the context of input text.

As technological advancements continue, newer and more innovative solutions have emerged. Particularly, recent generative artificial intelligence (AI) models offer a revolutionary approach to KM challenges. Global information technology (IT) and AI companies are releasing large language models (LLM) at a rapid pace and up-to-date representative models are ChatGPT and LLaMA introduced by OpenAI and Meta AI, respectively. Fine-tuning these AI models with domain-specific datasets enables a corporation to develop own conversational interface services [15]. It makes a large amount of technical knowledge readily accessible in a user-friendly question and answer (Q&A) format. These AI models represent significant advances over traditional database search tools. Fine-tuned AI models can understand natural language queries, making them more intuitive and user-friendly. They can interpret the intent behind the query, provide contextually relevant answers, and handle ambiguous or poorly structured questions. This not only improves the efficiency of information retrieval, but also improves the quality of the results, leading to a more informed decision-making process [16].

Leveraging AI models with in-house data can be an innovative solution, addressing the inherent challenges in KM by facilitating smoother and interactive knowledge acquisition processes. By integrating AI, organizations can streamline management processes and dramatically improve access to specialized knowledge, driving both learning and performance to new heights. Against this backdrop, this paper presents an LLM-driven solution that taps into the potential of these advanced models for in-house KM focusing on technical specification documents in engineering.

2. Research Background

2.1. Knowledge Management in Construction

KM is an integrated, systematic approach to identify, manage, and share an organization's knowledge [17]. It enables individuals to collaboratively create new knowledge and ultimately aiding in achieving the organization's objectives [18]. Effective KM is also critical in large-scale and complex construction projects. Globally, many engineering corporations worldwide are facing difficulties due to losing experienced professionals and the indispensable knowledge and skills they bring [19]. As many engineers approach retirement age, there is an evident lack of qualified junior engineers to take their places. Thus, it is necessary to transfer the tacit knowledge of seasoned professionals to organizational assets to replace senior engineers, managers, and experienced workers effectively. From a broader perspective, managing knowledge effectively ensures that the organization is utilizing all the resources available in the most effective manner to improve organizational performance [6]. However, many organizations face a lack of well-documented and integrated KM processes within their systems, especially in capturing and learning from past experiences [20]. To address these challenges, an effective KM system should incorporate methodologies that ensure continuous learning and knowledge sharing across the organization. A systematic approach to managing and sharing organizational knowledge helps in achieving organizational objectives efficiently and effectively, reducing the risk of knowledge loss in the complex environment [21]. In this context, the incorporation of innovative technologies becomes pivotal.

KM in the construction domain has been of interest to many researchers and practitioners. Previous KM research dominantly dealt with IT-based KM approach was performed from diverse perspectives, such as assessing the impact of KM [22], developing a KM platform [2,18,23], investigating KM strategy [4,9,10,24–26]. In the meantime, with the development of NLP technology, researchers have been interested in how to apply text analytics techniques from the perspective of document management and knowledge extraction. Rezgui (2006) proposed an ontology-based KM system that provides information retrieval service [18]. Based on the construction ontology expanded through industrial foundation classes (IFC). The proposed KM tool initially summarizes documents as index terms based on frequency. Then, the extracted index terms are matched against the construction ontology to retrieve relevant information with the user query. Salama and El-Gohary (2016) presented a machine learning model that classifies text, whether the given clause of a general contract belongs to a specific category or not [11]. In this study, they applied the bag of words for feature selection and the term frequency-inverse document frequency (TF-IDF) for feature weighting. Moon et al. introduced a named entity recognition (NER) model for automatic review of construction specifications [14]. They trained a neural network model to extract a word (or phrase) from specification documents and classify it into predefined categories: organization, action, element, standard, and reference. In the following study, Moon et al. proposed a web-based KM system that determines whether the given two provisions are relevant or not [1]. In this study, they used a semantic construction thesaurus developed by an ML-based word embedding technique.

Despite the application of existing KM tools and techniques has produced meaningful fruits, previous approaches have limitations. First, previous IT-based KM tools focused on information retrieval rather than information extraction. Keyword-based information retrieval searches for related documents based on the presence or absence of specific terms without considering the entire context of the text. Some previous KM tools applied ontology and NLP techniques to take into account the semantics of the text. Nevertheless, it required labor-intensive manual work to develop domain-specific ontology and training datasets for machine learning. In addition, it is difficult to consider newly emerging concepts that have not been used in the past, and there are many constraints to updating the developed ontology or NLP model.

2.2. Large Language Model

NLP is a branch of artificial intelligence enabling a computer to understand and interpret human language. There are various NLP tasks such as machine translation, question answering, and summarization. It involves the design and implementation of models, systems, and algorithms to solve practical problems using unstructured text data [27,28]. NLP has not only evolved as a theoretical concept but has also seen practical applications in diverse industries. With the increasing complexity of corporate technical documents, it is essential to leverage these advances for practical solutions. The construction domain is also responding to the development of NLP-based construction management technology, and the technology gap is gradually narrowing [29]. For instance, the automation of construction specification review through the application of NLP techniques has been previously studied and proven to be effective [1,30]. Also, the feasibility of a solution for finding information in long and complex policy documents difficult to navigate has been reviewed [28]. This advancement underscores the possibility of using NLP for automating complex tasks such as the management and interpretation of intricate technical specifications.

LLM is a pre-trained language model based on a vast amount of textual data using deep learning techniques. LLMs are characterized by their extensive scale and profound capabilities in language comprehension and generation [31]. Unlike their predecessors, these models are extensively trained on large textual datasets, enabling them to recognize complex linguistic patterns and establish connections between entities within language [31]. Such comprehensive training makes LLMs available to execute an array of language-related tasks to a remarkable degree of accuracy [32]. Employing advanced deep learning techniques, especially transformer architectures, these models process and produce language that can emulate human-like text across a diverse range of topics and styles [16]. The extensive parameter size of LLMs, often encompassing millions of parameters, empowers them to create coherent and contextually appropriate narratives, responses, and analyses. This capability is exceptionally beneficial in scenarios demanding an in-depth understanding of language, which includes, but is not limited to, conversational AI, content creation, and intricate information extraction endeavors.

Recent LLMs can be categorized into two types based on their source code availability: closed-source LLMs and open-source LLMs [33]. Closed-source LLMs are usually developed by global IT companies, and they provide API services for fine-tuning instead of providing source code publicly. Therefore, closed-source LLMs are restricted in customization because the developers allow limited functions for fine-tuning. Meanwhile, open-source LLMs provide more flexibility and controllability by opening their source code to the public. Based on the high level of customization, open-source LLMs can be utilized for specialized applications.

2.2.1. GPT

GPT-3 is a third-generation autoregressive language model that uses deep learning to produce human-like text [34]. It is a computational system designed to generate sequences of words, code, or other data, starting from a source input, called the prompt. The first generation of GPT in 2018 used 110 million learning parameters, and GPT-3 uses 175 billion parameters. GPT-3's architecture is based on the concept of transformers, a significant advancement in deep learning [35]. A transformer consists of two main parts: the encoder and the decoder. The encoder processes the input data, and the decoder generates the output. However, models like GPT focus only on the decoder part for language generation tasks. GPT models are first pre-trained on a massive corpus of text data. This stage involves learning language patterns, grammar, and context without any specific task in mind. GPT-3 was pre-trained on an unlabeled dataset that is made up of texts, such as Wikipedia and many other sites. Although there are copyright concerns with the text data used for training [36], LLMs have been able to reach AI human-like language abilities by using a large amount of text data available [31]. After pre-training, GPT models can be fine-tuned for specific tasks like translation, question-answering, or text generation. This involves additional training with task-specific dataset. However, it requires relatively small amount of additional training data for fine-tuning compared to the language models before GPT.

2.2.2. LLaMA

LLaMA is an open-source LLM released by Meta [37]. Like other recent LLMs, the foundation of LLaMA is also based on the transformer architecture [35]. Both GPT and LLaMA are designed to understand the context of human language and generate natural language by trained with vast amounts of text data. However, they differ in the size and diversity of the training data, how recent the pretraining data is, and the specific techniques used to train the models [37]. Moreover, the approach to sharing LLaMA offers more open access or different licensing, which can influence how researchers and developers can use the technology to resolve domain-specific problems. LLaMA2 is a modified version of the original LLaMA. It upgraded the original LLaMA by expanding and robustly cleaning the training dataset, increasing the context length, and using grouped-query attention for better inference scalability [38]. As a result, LLaMA2 recorded significant performance improvements in overall NLP tasks compared to other open-source LLMs [33].

2.3. Fine-Tuning LLMs

Fine-tuning is a process in which a pre-trained language model is further trained on a specific dataset to perform user-defined tasks more effectively. It aims to transfer the knowledge and capabilities learned during the initial pre-training phase to a narrower, domain-specific task. Fine-tuning allows a user to leverage an LLM's natural language understanding and generation abilities for specific applications. Fine-tuning LLMs has distinct differences from traditional fine-tuning methods due to the complexity and size of LLMs [39]. The concept of fine-tuning is to enable the model to adjust its parameters to better fit the task-specific data while retaining the general language understanding it acquired during pre-training. Traditional fine-tuning methods involve updating all or a substantial portion of the parameters of a pre-trained model. However, recently introduced LLMs have been much larger in terms of the number of parameters, which makes fine-tuning them more resource-intensive. Updating all the parameters of LLMs requires not only significant computational resources but also substantial memory and storage. Against this backdrop, fine-tuning LLMs has been studied in the light of a more resource-efficient approach, called parameter-efficient fine-tuning (PEFT) methods.

The primary concept of PEFT is to fine-tune only a small subset of the LLM's parameters during the fine-tuning process [40]. This approach contrasts with traditional fine-tuning methods where all the parameters of a pre-trained model are updated. PEFT is useful for adapting large models to specific tasks or datasets without extensive computational cost and time required to retrain the entire model. PEFT allows for more flexible adaptation to domain-specific tasks, as the core pre-trained model remains largely unchanged and only small, task-specific adjustments are made.

2.4. State of the Art and Research Gaps

Recent LLMs are expected to shift the working paradigm in most industries. Both public and private sectors attempt to leverage generative AI models for productivity improvement. AI models can be divided into two types: close- and open-source models. It is difficult to say which one is better than the other. Closed-source models, which are usually commercial models, do not share detailed information such as model architecture or parameter size. On the other hand, open-source models make their source code available to everyone. This difference affects how users choose and use a language model for their needs. When utilizing closed-source models, users have to send their data to the service provider for fine-tuning, but the process and how to fine-tune the model's performance are not disclosed. Moreover, although service providers guarantee that data provided by users will not be used anywhere else, ensuring security and confidentiality remains a concern. These limit what users can do and pose challenges in securing the applicability of models. Therefore, practitioners prefer to use open-sourced models to develop their own AI models.

This study aims to extend the utilization of LLM in the corporate context. In detail, this study proposes an LLM-driven KM tool that makes company standard documents easily accessible and understandable, not only for professionals but also for those who are new to the field. By comparing closed- and open-source LLMs fine-tuned with engineering specifications, this study attempts to analyze which type of model is better for solving domain-specific problems.

NLP can be used to automate corporate tasks in the context of KM as follows. First, it can be used to classify documents into different categories and to retrieve relevant documents based on user queries. Second, key knowledge can be extracted from documents such as concepts, relationships, and events. Third, NLP can be utilized to develop a chatbot that can answer user questions in a comprehensive and informative way. Among the aforementioned applications of NLP techniques, this study aims to explore an LLM-driven KM interface based on a question-answering chatbot capable of extracting relevant information from in-house engineering documents.

3. Research Methodology

This study presents LLM-based Q&A models for technical specifications focusing on the pressure vessel. Figure 1 shows the research process of this study. The authors collected the technical specifications of three companies and preprocessed them to develop Q&A datasets. During the data preprocessing, this study expanded the original dataset using paraphrasing and cross-matching methods to enable LLMs to learn more diverse representations of specifications. Then, this study fine-tuned two representative LLMs. One is the Davinci-002 model, a closed-source LLM introduced by OpenAI. The other is the LLaMA2-13B model, an open-source LLM introduced by Meta. The authors evaluated the fine-tuning results of the two LLMs based on the gold standard developed by experts in this field.

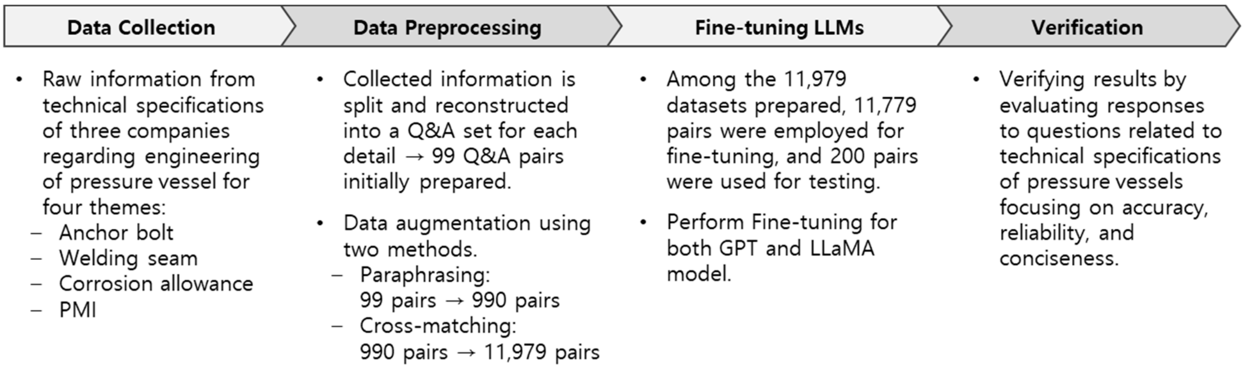

Data Collection
Data Preprocessing
Fine-tuning LLMs
Verification
• Raw information from technical specifications of three companies regarding engineering of pressure vessel for four themes:
– Anchor bolt
– Welding seam
– Corrosion allowance
– PMI
• Collected information is split and reconstructed into a Q&A set for each detail → 99 Q&A pairs initially prepared.
• Data augmentation using two methods.
– Paraphrasing: 99 pairs → 990 pairs
– Cross-matching: 990 pairs → 11,979 pairs
• Among the 11,979 datasets prepared, 11,779 pairs were employed for fine-tuning, and 200 pairs were used for testing.
• Perform Fine-tuning for both GPT and LLaMA model.
• Verifying results by evaluating responses to questions related to technical specifications of pressure vessels focusing on accuracy, reliability, and conciseness.

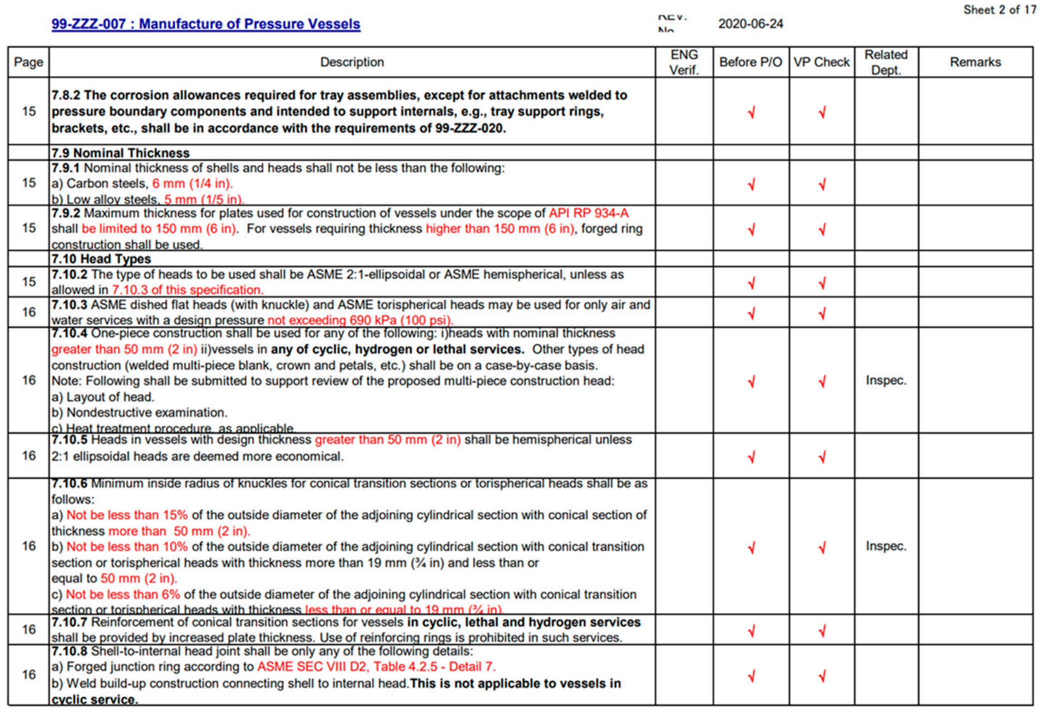
Figure 1. Research process.

3.1. Data Collection

In many mega-projects, it is critical to ensure that technical specification documents are carefully reviewed and consistently applied. To achieve this, many companies use a "Technical Specification Summary Sheet" at the beginning of their engineering projects. This sheet serves as a comprehensive checklist, highlighting applicability, key considerations, and essential cautionary notes in a systematically tabulated format. Figure 2 provides an illustration of this summary sheet as it is typically used in engineering practice.

For the model development, this study selected the pressure vessel as the subject due to its ubiquity and foundational role in plant engineering. Furthermore, every company typically maintains technical specifications for such equipment. This study constructed a foundational dataset derived from technical specifications related to pressure vessel engineering. The authors collected the specifications from three companies. The specifications used in this study are internal data owned by the company to which one of the authors is affiliated. Here, the authors removed and modified all information within the training dataset that could represent the specific companies by converting the names of three companies as AAA, BBB, and CCC because of security.

To bring diversity and structure to the specifications, this study selected four themes considering several intents and their frequent mentions in the documents: welding seam, corrosion allowance, anchor bolt, and positive material identification (Table 1). These themes subsequently formed the foundation of the Q&A dataset for model fine-tuning. Choosing specifications from multiple companies for a single type of equipment, such as a pressure vessel, ensured a richer and more diverse dataset. This choice becomes especially relevant when considering that similar content might be present across a company's specification, international codes, and regulatory body requirements. However, it is also important to note that in real-world scenarios, a company might not have multiple specifications for a single equipment type. Nonetheless, sourcing from multiple entities not only provided a broader range of questions and answers but also tested the model's ability to distinguish between closely related content.



Sheet 2 of 17

99-ZZZ-007 : Manufacture of Pressure Vessels REV. No. 2020-06-24

Page	Description	ENG Verif.	Before P/O	VP Check	Related Dept.	Remarks
15	**7.8.2 The corrosion allowances required for tray assemblies, except for attachments welded to pressure boundary components and intended to support internals, e.g., tray support rings, brackets, etc., shall be in accordance with the requirements of 99-ZZZ-020.**		√	√		
	7.9 Nominal Thickness					
15	**7.9.1** Nominal thickness of shells and heads shall not be less than the following: a) Carbon steels, 6 mm (1/4 in). b) Low alloy steels, 5 mm (1/5 in).		√	√		
15	**7.9.2** Maximum thickness for plates used for construction of vessels under the scope of API RP 934-A shall be limited to 150 mm (6 in). For vessels requiring thickness higher than 150 mm (6 in), forged ring construction shall be used.		√	√		
	7.10 Head Types					
15	**7.10.2** The type of heads to be used shall be ASME 2:1-ellipsoidal or ASME hemispherical, unless as allowed in 7.10.3 of this specification.		√	√		
16	**7.10.3** ASME dished flat heads (with knuckle) and ASME torispherical heads may be used for only air and water services with a design pressure not exceeding 690 kPa (100 psi).		√	√		
16	**7.10.4** One-piece construction shall be used for any of the following: i)heads with nominal thickness greater than 50 mm (2 in) ii)vessels in **any of cyclic, hydrogen or lethal services.** Other types of head construction (welded multi-piece blank, crown and petals, etc.) shall be on a case-by-case basis. Note: Following shall be submitted to support review of the proposed multi-piece construction head: a) Layout of head. b) Nondestructive examination. c) Heat treatment procedure, as applicable.		√	√	Inspec.	
16	**7.10.5** Heads in vessels with design thickness greater than 50 mm (2 in) shall be hemispherical unless 2:1 ellipsoidal heads are deemed more economical.		√	√		
16	**7.10.6** Minimum inside radius of knuckles for conical transition sections or torispherical heads shall be as follows: a) Not be less than 15% of the outside diameter of the adjoining cylindrical section with conical section of thickness more than 50 mm (2 in). b) Not be less than 10% of the outside diameter of the adjoining cylindrical section with conical transition section or torispherical heads with thickness more than 19 mm (¾ in) and less than or equal to 50 mm (2 in). c) Not be less than 6% of the outside diameter of the adjoining cylindrical section with conical transition section or torispherical heads with thickness less than or equal to 19 mm (¾ in).		√	√	Inspec.	
16	**7.10.7** Reinforcement of conical transition sections for vessels **in cyclic, lethal and hydrogen services** shall be provided by increased plate thickness. Use of reinforcing rings is prohibited in such services.		√	√		
16	**7.10.8** Shell-to-internal head joint shall be only any of the following details: a) Forged junction ring according to ASME SEC VIII D2, Table 4.2.5 - Detail 7. b) Weld build-up construction connecting shell to internal head.**This is not applicable to vessels in cyclic service.**		√	√		

Figure 2. Sample technical specification summary sheet.

Table 1. Specification types.

Type	Check Point
Anchor bolt	• Does it accurately provide design-related numerical values (e.g., stress, material codes) without confusion or error? • Check hallucination issue.
Corrosion allowance	• Does it distinguish and respond to subtle differences in textual content per specification?
Welding seam	• Does it distinguish and respond to subtle differences in content (text, numbers) per specification? • Check hallucination issue.
Positive material identification	• Does it reference related document numbers correctly? • Check hallucination issue.

3.2. Data Preprocessing

The raw information is split and reorganized into a Q&A form for each detail. After collecting the raw technical specifications, a comprehensive preprocessing step began. The raw information was meticulously segmented, analyzed, and reorganized into 33 intuitive question–answer pairs. This format was chosen to simulate potential queries that might arise in real-world scenarios, ensuring that the model can be practically applied to support field experts. The reconstruction of the set was done manually by engineers with extensive practical experience.

Since the length and composition of each specification was slightly different, the number of Q&A sets that could be extracted varied, as shown in Table 2. In the case of companies where the content of the theme was divided into separate specifications, there were cases where the number of Q&A sets that could be extracted was small. From these initial 33 question–answer pairs, questions from all companies were combined for each type (theme) to ensure consistent data application across each company. For instance, questions extracted from AAA were also asked to BBB and CCC, and questions extracted from BBB were also asked to AAA and CCC, thereby balancing the questions. The answers are different depending on the specifications of each company. As a result, the initial 33 Q&A pairs increased to up to 99 original pairs.

Table 2. The number of extracted Q&A pairs.

Specification Type	AAA	BBB	CCC	Total
Anchor bolt	3	1	5	9
Corrosion allowance	4	2	3	9
Welding seam	1	7	3	11
Positive material identification	2	1	1	4
Total	10	11	12	33

This study applied two-step data augmentation techniques to diversify our initial dataset. At first, the authors paraphrased the original Q&A dataset in order to enrich the dataset with different syntactic structures. When fine-tuning an LLM, paraphrasing is effective strategy to enable a model to learn diverse representations [41]. Thus, this study paraphrased each of the 99 questions and 99 answers 10 times, generating an additional 990 pairs. As a result, a total of 1089 Q&A pairs were prepared including original and paraphrased Q&A pairs. Then, 11 questions and 11 answers from each of the 99 original Q&A pairs were cross-matched to produce unique combinations, yielding a final set of 11,979 pairs ($11 \times 11 \times 99$). This approach further strengthened the contextual associations between the questions and answers.

After confirming the diversity and comprehensiveness of the dataset, the authors split all the data into training and test datasets using stratified sampling. When splitting the datasets, this study considered that the number of companies and specification types in each dataset should be similar. Here, the number of test dataset pairs in this study was intentionally limited to 200. Considering the given spectrum of data variations, this count was deemed sufficient to provide a representative sample for human assessment, encompassing 2% of the total dataset of 11,979 pairs. The remaining 11,779 pairs, constituting 98% of the dataset, were allocated for fine-tuning.

3.3. Fine-Tuning LLMs

This study employed representative LLMs from each of the closed-source and open-source LLMs considering the accessibility and trialability of each LLM for the purpose of comparison, namely, Davinci-002 model, which is one of GPT series, introduced by OpenAI and LLaMA2-13B model released by Meta. The Davinci-002 model was fine-tuned through the API service provided by OpenAI and the LLaMA2 model by using the Quantized Low-Rank Adaptation (QLoRA) method [42]. One LLM this study fine-tuned is the Davince-002 model. Because GPT series are closed-source LLMs, the authors used the API service provided by OpenAI for fine-tuning. The GPT-3 tuning procedure requires specificity. Data sets should conform to a prompt-completion architecture, preferably in JSONL format. This structure makes capturing the correlation between the prompt and its corresponding completion easier for the model. A prompt is a query that a user provides to elicit a response from the LLM. The LLM utilizes the prompt to produce a response based on pretrained data. The authors prepared prompt-completion pair datasets in JSONL format in accordance with the instructions guided by OpenAI [43]. Each prompt-completion pair should be in a dictionary format as shown in Figure 3. Here, '\n\n###\n\n' in a prompt and '#END' in a completion indicate the end of a prompt and completion, respectively. After preparing and uploading the dataset, the fine-tuning process is performed automatically through the API.

Another LLM this study fine-tuned is the LLaMA-13B model. Traditional fine-tuning of pre-trained language models commonly update all parameters based on additional training datasets. However, researchers have been investigating how to fine-tune a pre-trained language model efficiently as the size of language models increases. PEFT aims to solve specific problems in a short period with low resources and computation by further training a small number of model parameters. Various PEFT methods have been introduced recently such as Adapters [44], Low-Rank Adaptation of Large Language Models (LoRA) [45], and $(IA)^3$ [46]. QLoRA is one of the representative PEFT methods recently employed, and this study fine-tuned the LLaMA-13B model using the QLoRA method. The QLoRA method combines two core PEFT techniques: quantization and LoRA. Quantization is a method that converts a data type of the original transformer model to the 4-bit NormalFloat data type to reduce memory requirements for fine-tuning [42]. Although there is a trade-off between precision loss and memory efficiency, leveraging the quantization is worthwhile because it enables fine-tuning a larger size of LLM with lower resources. LoRA is one of the PEFT techniques used in fine-tuning LLMs. It offers an efficient and effective approach to modifying pre-trained language models for specific tasks. In order to reduce computing costs, the LoRA method trains an additional small subset of the parameters while freezing all the parameters of an original LLM [45]. LoRA decomposes the weight matrices of the original transformer model into lower-rank matrices, which are trainable. Thus, it makes fine-tuning process more parameter-efficient compared to full model fine-tuning.

```
{“prompt”: “What is the minimum nominal diameter for anchor bolts according to the BBB company
specification?\n\n###\n\n”,
 “completion”: “The BBB company specification states that anchor bolts should not be less than
19 mm (3/4 in) in minimum nominal diameter.#END”}
{“prompt”: “What does the AAA company specification say about welding seam
information?\n\n###\n\n”,
 “completion”: “Based on the specification from the AAA company, If shell courses have
different thicknesses, the inside diameter should remain constant unless specified otherwise.
The minimum distance between two longitudinal seams in one course must be either 200 mm (8
inches) or five times the wall thickness, whichever is larger. This distance is measured
between the toes of each weld. There should be a minimum distance between staggered
longitudinal seams of two adjacent courses, the minimum distance should be 200 mm (8 inches) or
five times the wall thickness, whichever is greater. If maintaining the minimum distance
between seams is unachievable, the last 300 mm (12 inches) of the adjacent longitudinal and the
intersected circumferential seams must undergo 100% examination in accordance with sections
(6.3) or (6.4), whichever is applicable.#END”}
{“prompt”: “What does the CCC company specification state about the corrosion
allowance?\n\n###\n\n”,
 “completion”: “The CCC company specification provided states the following regarding corrosion
allowance:. Section 2.7.1: In vessels or sections of vessels made of high alloy, or those that
are alloy lined or clad, the surfaces of pressure parts exposed to contained media, and all
surfaces of non-pressure parts which are non-removable and exposed to the contained media,
should either be fabricated of or protected by the alloy material specified for the vessel.
Section 2.7.2: In unlined vessels or sections of vessels, corrosion allowance should be added
to each exposed surface of all surfaces of pressure parts exposed to the contained media and to
non-pressure parts welded directly to the pressure boundary and those parts which are otherwise
non-removable and exposed to the contained media. However, corrosion allowance shall not be
added to each surface of replaceable non-pressure parts. Section 2.7.3: The minimum corroded
thickness of non-removable parts, parts welded directly to the vessel, and the throat of fillet
welds shall be 3 mm (1/8 in).#END”}
{“prompt”: “Where can one find the PMI requirements according to the CCC company
specification?\n\n###\n\n”,
 “completion”: “According to the CCC company's specification, the PMI requirements can be found
in the “Purchaser Technical Standard 2-1CCC” document.#END”}
```

Figure 3. Example of prompt-completion pairs in JSONL format.

This study fine-tuned the LLaMA2-13B model in a local development environment while the Davinci-002 was fine-tuned through API service provided by OpenAI. The local development environment was Ubuntu 20.04 and Python version 3.11.5 with an AMD Ryzen 9-5950X CPU and a GeForce RTX-3090 Ti GPU. In addition, this study used the same hyperparameters used in QLoRA fine-tuning experiments performed by Dettmers et al. [42] in which the QLoRA was first introduced (Table 3).

Table 3. QLoRA Hyperparameter setting for LLaMA2-13B fine-tuning.

Hyperparameter	Value
LoRA attention dimension	64
Alpha parameter for scaling	16
Dropout probability	0.1
Learning rate	2×10^{-4}
Batch size	16

When utilizing an LLM, the 'Temperature' parameter affects the output of the model. It is an important parameter when it comes to controlling the randomness of the text generated by LLMs [47]. The ideal temperature setting can vary depending on the goal of the query and how much variation or creativity a user desires in the responses. It can be set from 0 to 2. When the temperature is set to 0, the model generates the most probable word or phrase in its potential outputs, which leads to the output being very deterministic and giving predictable answers. An LLM generates diverse and creative responses when the temperature increases. The randomness of the output becomes high, which means the model is much more likely to produce unexpected or non-traditional answers. Thus, this study sets the temperature as 0 because it is necessary to provide an answer based on as much fine-tuned data as possible.

3.4. Gold Standard for Evaluation

This study evaluated the effectiveness of fine-tuning using the 200 prompt-completion pairs that were not used in fine-tuning. Completions generated by the fine-tuned models were evaluated and scored by experts with practical knowledge in the field. This study established an evaluation guideline considering accuracy, reliability, and conciseness of responses. The accuracy of responses evaluates how precisely the model's responses align with the correct or expected answers. It measures the degree to which the answers are factually correct, contextually appropriate, and relevant to the questions asked. Accuracy is critical as it reflects the model's understanding of the query and its ability to provide correct information. The reliability of responses assesses the consistency and dependability of the model's answers over time and across various question types. This criterion checks whether the model consistently provides correct and relevant answers, maintaining a standard level of performance regardless of the complexity or nature of the queries. High reliability ensures that users can trust the model's responses in different contexts and scenarios. Lastly, the conciseness of responses: Conciseness measures the ability of the model to provide clear and brief answers, without unnecessary elaboration or information that is not directly relevant to the question. This criterion ensures that the model's responses are to the point, making them more accessible and easier to understand for users. It is a balance between providing enough information to satisfactorily answer the question and keeping the response succinct and focused. Table 4 represents the evaluation guideline for establishing the gold standard of LLMs' responses, and Table 5 shows evaluation examples of representative ratings of 1, 3, and 5.

Table 4. Evaluation guideline of the responses of fine-tuned LLMs.

Rating	Description
5	The highest score is given if the question is answered correctly. Even if your answer is longer than necessary or contains ambiguous words, it is awarded this score compared to the best answer because it fundamentally solves the problem accurately.
4	This score is awarded if the answer is related to the question but contains minor errors or content that is irrelevant to the question.

Table 4. *Cont.*

Rating	Description
3	Give this score to answers that convey the overall meaning and include keywords, but have unnecessary additions or are missing important information. The core message should be delivered without major inaccuracies.
2	Answers that are related to the question but are inaccurate or have low reliability due to incorrect information.
1	The lowest score is given to answers that are completely irrelevant or whose meaning is not understandable.

Table 5. Representative evaluation examples.

Rating	Class	Example
5	Prompt	Based on the CCC company specifications, what is the required offset for longitudinal seams in adjacent shell rings?
	Exemplary answer	According to the CCC company specifications, adjacent shell rings should be offset by a minimum of four times the plate thickness or 150 mm (6 in).
	LLM's response	A longitudinal seam in adjacent shell courses (or rings) should be offset by at least four times the plate thickness or 150 mm (6 in), whichever is greater, according to CCC company specifications.
3	Prompt	According to the CCC company specifications, how much offset is required for longitudinal seams in adjacent shell rings?
	Exemplary answer	CCC specifications require longitudinal seams in adjacent shell rings to be offset by a minimum of four times the plate thickness or 150 mm (6 in).
	LLM's response	A longitudinal seam in adjacent courses must clear the outside diameter of the vessel by four times the thickness of the plate, or 150 mm (6 in), whichever is greater, according to the CCC company's specification. The last course of the higher side and five times the thickness of the plate should be subjected to 100% examination in accordance with section (6.3) or (6.4). The minimum distance between the last longitudinal seam of the adjacent nozzle and the nozzle reinforcing pad is 50 mm (2 in), or three times the wall thickness, whichever is greater. Overlap of longitudinal seams in adjacent vessels . . ., longitudinal seams in adjacent vessels or sections of vessels should not be overlapped by more than 150 mm (6 inches) or 5
1	Prompt	What is the design specification for carbon steel anchor bolting according to the AAA company?
	Exemplary answer	It is necessary to search for other related specifications.
	LLM's response	The material specification, 22

4. Results

The overall evaluation results for the fine-tuned Davinci-002 and LLaMA2-13B models are presented in Figure 4. LLaMA2-13B model outperformed the Davinci-002 model with a higher average score and a higher frequency of the top score. The average scores of the LLaMA2-13B and Davinci-002 models were 3.56 and 2.82, respectively. The median of the LLaMA2-13B model was 5, which indicates that the LLaMA2-13B model correctly responded to more than half of the entire questions in the test dataset. Meanwhile, the median of the Davinci-002 model was 2.5, which indicates that half of the responses produced by the Davinci-002 model were evaluated to be inadequate.

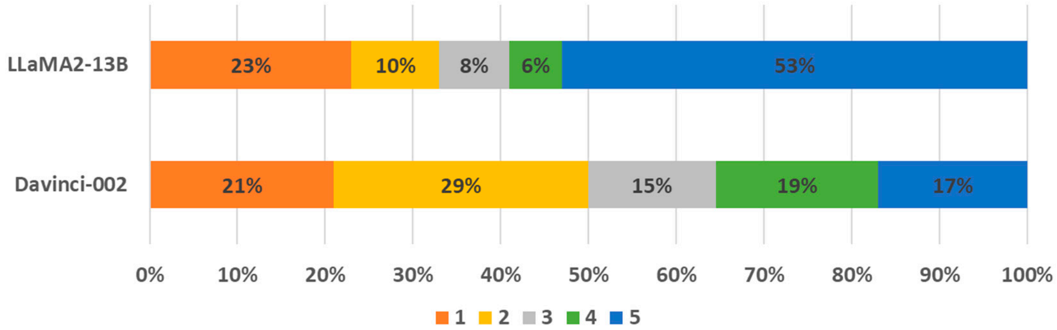


Figure 4. Overall fine-tuned LLMs performance.

The frequency of the highest score was high for the LLaMA2-13B model for all specification types, and the frequency of the lowest score was low for the Davinci-002 model, all showing a similar pattern to the overall results. Among the four specification types, three types except the corrosion allowance type showed results in which the LLaMA2-13B model had the highest score of more than 50% (Table 6).

Table 6. Fine-tuned LLMs performance by specification type.

Rating	**LLaMA2-13B**					**Davinci-002**				
	5	**4**	**3**	**2**	**1**	**5**	**4**	**3**	**2**	**1**
Anchor bolt	60%	4%	4%	7%	25%	7%	13%	22%	13%	45%
Corrosion allowance	39%	11%	11%	13%	26%	18%	18%	19%	37%	8%
Welding seam	51%	4%	9%	9%	27%	15%	18%	4%	42%	22%
Positive material identification	75%	4%	7%	11%	4%	39%	32%	11%	18%	0%

The authors verified the results by classifying responses into four categories according to their characteristics: 'Search for other related specification', 'Explicit specification requirement', 'How is it determined?' and 'Refer to [another specification no.]'. 'Search for other related specification' category indicates that the current specification does not contain the information requested by the questioner and that additional specifications should be consulted to obtain relevant details. 'Explicit specification requirement' category is when a specification contains clear, direct statements, instructions, or numerical values that detail a specific requirement or standard to be followed. 'How is it determined?' refers to the method or process for establishing specific requirements. This may refer to the process for determining appropriate specifications, the decision maker or responsible entity, or a reference to a specific data sheet. Lastly, 'Refer to [another specification no.]' specifies the exact document number that the reader should consult another document to find detailed information about a particular aspect of the specification that is not included in the current specification.

Table 7 shows the performance of the LLaMA2-13B and Davinci-002 models by response type. The Davinci-002 model showed poor results in all response types compared to the LLaMA2-13B model. Especially, the LLaMA2-13 model showed significant performance in responding to the 'Search for other related specification' category with 80% of perfect answers, followed by 'Refer to [another specification no.]' with 68% of correct answers. Meanwhile, both Davinci-002 and LLaMA2-13B models showed poor performance in extracting 'Explicit specification requirement'. Although the responses of the LLaMA2-13B model were better than that of the Davinci-002 model overall, it is noteworthy that the LLaMA2-13B model was rated with a higher percentage of lowest scores than the Davinci-002 model in three response types except the 'Search for other related specification'. In-depth analysis of the results will be discussed in the following section.

Table 7. Fine-tuned LLMs performance by response type.

Rating	LLaMA2-13B					Davinci-002				
	5	4	3	2	1	5	4	3	2	1
Search for other related specification	80%	3%	5%	5%	8%	14%	8%	9%	18%	51%
Explicit specification requirement	29%	6%	11%	13%	41%	11%	21%	19%	42%	7%
How is it determined?	42%	21%	11%	11%	16%	26%	26%	11%	32%	5%
Refer to [another specification no.]	68%	3%	6%	13%	10%	35%	29%	16%	13%	6%

This study also compared the performance of the LLaMA2-13B and Davinci-002 models by prompt-completion type. The authors classified the prompt-completion pairs into four categories. Type 1 indicates that only one of the three companies has appropriate specifications related to the user query, and the other two companies do not have relevant content in their specifications. The type 2 prompt-completion pairs have ambiguous content to distinguish each company's specification. Type 3 represents that each company's specification contains content that distinguishes it from the user query, but the content is vague. Finally, the type 4 prompt-completion pairs refer to the specifications of all three companies and contain numerical values and clear content.

In all types of prompt-completion pairs, the LLaMA2-13B model recorded a higher average score as shown in Table 8. In detail, the result of the LLaMA2-13B model tends to be polarized into the highest and lowest scores. The LLaMA2-13B model responded well to the user query, at least twice as much as the Davinci-002 model. However, the result of the LLaMA2-13B model recorded lower scores in the middle ratings of responses than the Davinci-002 model. In the case of type 3, 18% of the Davinci-002 model's responses received the lowest score, while 32% of the LLaMA2-13B model's responses received the lowest score.

Table 8. Fine-tuned LLMs performance by prompt-completion type.

Rating	LLaMA2-13B					Davinci-002				
	5	4	3	2	1	5	4	3	2	1
Type 4	54%	0%	4%	10%	32%	18%	14%	12%	38%	18%
Type 3	54%	12%	10%	12%	12%	26%	26%	18%	30%	0%
Type 2	42%	6%	12%	16%	24%	14%	8%	12%	44%	22%
Type 1	62%	6%	6%	2%	24%	10%	26%	16%	4%	44%

5. Discussion and Implications

In the existing engineering practice, keyword-based searches and manual document reviews such as specification summary sheets were mainly conducted. These manual tasks require a lot of time and effort from many people, and if a technical problem or human error occurs in the document managing, the result will be difficult to recover. However, the method suggested in this paper focuses on understanding natural language queries and searching for information faster and more efficiently than before. Further, it showed the idea of retaining and utilizing the history of user questions and model answers as in-house knowledge by reusing them as training datasets. It shows strength in automatically maintaining, transmitting, and sharing expert knowledge along with efficient knowledge search. Moreover, this knowledge is automatically maintained as the company's knowledge assets. Since personal memory has limitations, this automated procedure will greatly help internal KM. Moreover, the LLM-based KM presented in the study emphasizes natural language comprehension ability to identify query intentions and generate appropriate responses without the need for accurate keyword matching. This allows users to access information more intuitively and efficiently.

The results of this study suggest the potential applicability of the proposed approach. To the best of our knowledge, there are no examples of such an approach being used for navigating and managing technical specifications. It is shown that the GPT-3 model, once fine-tuned, might be able to deliver answers based on the content present in the selected dataset. The digitization and structuring of technical documents into an easy-to-navigate chatbot could potentially streamline the process for new employees who need to understand these documents. Additionally, the transition of a company's technical know-how into a database might serve as a defensive strategy against the possible loss of institutional knowledge. It acts as a tool to bring people together and enhance communication, and allows the organized storage and transfer of unstructured thoughts and notes, etc. [28]. Nonetheless, it should be noted that the AI model's ability to provide accurate answers is directly tied to the information included in the dataset. Thus, it is essential to continuously update and expand the content to maintain the relevance of the responses.

Although the fine-tuned model occasionally introduced extraneous details into the responses, the study found that engineers could easily identify and ignore irrelevant or redundant information. This observation suggests that the presence of additional, non-critical details in the model outputs does not significantly affect their usefulness in practical scenarios. Experienced engineers could easily identify and filter out irrelevant or redundant information added by the fine-tuned model. This observation suggests that the impact of such additional details is not significantly detrimental. It highlights the ongoing importance of expert interpretation and judgment when utilizing AI-based systems in real engineering environments.

5.1. Discussion on the Results by Specification Type

Analysis based on the intended use of the data shows clear strengths and weaknesses in each model. The LLaMA2-13B model generally outperformed the Davinci-002 model, particularly in accuracy and handling complex content. However, the Davinci-002 model's lower frequency of receiving the lowest scores in certain categories indicates areas where it maintains robust performance.

- Anchor bolt (Design-Related Numerical Values): The Davinci-002 model scored an average of 2.26, while the LLaMA2-13B model scored 3.86. This indicates the LLaMA2-13B model's superior ability in accurately providing design-related numerical values.
- Corrosion allowance (Textual Content Accuracy): The average score for the LLaMA2-13B model was higher at 3.16 compared to 2.88 for the Davinci-002 model. However, it is noteworthy that the Davinci-002 model received fewer lowest scores (1 point)-5 times compared to 16 times for the LLaMA2-13B model.
- Welding seam (Text + Number Contents): Here, the Davinci-002 model scored an average of 2.64, while the LLaMA2-13B model scored 3.64, indicating better performance of the LLaMA2-13B model in handling combined text and numeric content.
- Positive material identification (Referencing Accuracy): Both models scored high (Davinci-002 model: 3.94, LLaMA2-13B model: 4.36). The high scores in this category can be attributed to the relatively simpler and clearer reference document numbers used in training, which leave little room for creative interpretation by the generative LLMs.

Hallucination, in the context of artificial intelligence, refers to the phenomenon where an AI model produces plausible but incorrect answers or errors in information processing, similar to illusions or delusions. This can occur when the AI model produces results based on information that is not present in its training data, leading to results that may be coherent but factually incorrect, or even biased and of low relevance. Hallucinations in AI are often observed when the available data is scarce or incomplete. Generative LLMs can hallucinate, generating factually incorrect or unrealistic information. This issue, which arises from limitations in training data, inherent design, prompt ambiguity, and lack of real-world understanding, is critical in evaluating model responses. The hallucination tendency of the fine-tuned completion for a category is judged by the number of times the scores 1 and 2 were received. This is because if the post fine-tuning models include content other than 'Search for other related specification,' they will receive a low score. Among a total of 45 datasets corresponding to 'Search for other related specifications,' the Davinci-002 model recorded a score of 2 or less 18 times in the 'Anchor bolt' type, 13 times in the 'Corrosion allowance' type, and 14 times in the 'Welding seam' type. And in the case of the LLaMA2-13B model, the scores 1, 5, and 2 were observed in the order of 'Anchor bolt,' 'Corrosion allowance,' and 'Welding seam,', respectively. Thus, the hallucination tendency can be judged to be greater in the Davinci-002 model than in the LLaMA2-13B model.

5.2. Discussion on the Results by Response Type

By examining several key discussion points in the scoring system, ranging from the lowest score of 1 to the highest score of 5, the authors categorize scores below 2 as lower scores indicating impracticality for practical use, and scores above 4 as higher scores indicating practical utility. This analysis confirms that the fine-tuning performance of the LLaMA2-13B model is superior to that of the Davinci-002 model in most cases. However, the relatively higher performance of the Davinci-002 model in the 'Explicit specification requirement' category is noteworthy. These results provide insight into how each model responds to specific types of queries and can serve as important guidelines for future research and development directions.

- In the 'Search for other related specification' category, out of 65 fine-tuned completions, the Davinci-002 model scored above 4 in only 14 instances, while the LLaMA2-13B model did so in 54 instances. This significant difference suggests that the Davinci-002 model tends to not acknowledge its ignorance about content not included in its training data, a phenomenon known as hallucination. On the contrary, the LLaMA2-13B model showed a higher frequency of accurately indicating its lack of knowledge when asked about content absent in its fine-tuned material. This is further evidenced by the number of scores below 2 in this category, where the Davinci-002 model scored 45 and the LLaMA2-13B model only 8.
- Examining the highest score of 5, the LLaMA2-13B model scored significantly higher frequencies of 5 (ranging from 2 to 6 times higher) in all categories except 'How is it determined?' category. Even in this category, though the frequency difference was marginal, the LLaMA2-13B model still outscored the Davinci-002 model in terms of the number of the score 5 received.
- Looking at the lowest score of 1, in the 'Search for other related specification' category, the LLaMA2-13B model had significantly fewer instances, while it was the Davinci-002 model that had significantly fewer 1s in the 'Explicit specification requirement' category. Most of the results indicated a superior fine-tuning performance of the LLaMA2-13B model; however, interestingly, the Davinci-002 model received significantly fewer 1s in the "Explicit specification requirement" category and thus had a slightly higher average score (Davinci-002 model: 2.86, LLaMA2-13B model: 2.69).

Data privacy and security are key concerns that must be addressed when developing LLM-based organizational KM tools for industrial use. The fine-tuning process involves the use of potentially sensitive corporate information, and it is essential that strict measures are implemented to prevent unauthorized access or data breaches. Although it requires additional processes and deeper knowledge, the use of open-source models offers advantages in developing a domain-specific model that is not only more accurate but also more cost-effective.

6. Conclusions

This study suggests that the GPT-3 and LLaMA2 models, once fine-tuned, might be capable of delivering answers based on the content in the selected dataset. This is a critical consideration for optimizing the sharing of proprietary knowledge within an enterprise or organization. Enhancing KM with digital technology and artificial intelligence can offer additional benefits in collating knowledge in the engineering and construction industry. Digitizing and structuring technical documents into an easily navigable interface could potentially aid new employees in understanding these documents and help mitigate the risk of losing essential knowledge when key staff depart.

Current engineering practices heavily rely on manual searches and document reviews, which are time-consuming and prone to errors. Against this backdrop, this paper presented a new approach that automatically retrieves information faster and more efficiently than the traditional approach. The proposed LLM-based model is trained using past user questions and responses to improve KM. The result of this study highlights the use of cutting-edge AI models to understand user queries better and provide relevant responses, making information access easier and quicker. Consequently, the proposed method aims to automate KM, making it easier to handle and share information within the company. It opens new pathways to enhance organizational efficiency and productivity by delivering immediate answers to user queries.

However, several limitations and improvement points must be noted. First, as previously mentioned, the model's ability to provide accurate answers is intrinsically limited to the content and scope of the dataset. This necessitates a commitment to regular updates and dataset expansion. To create a KM tool that handles entire tasks, it is required to gather knowledge and knowhow of individuals to transfer personal knowledge into organizational assets. Second, technological constraints, especially the current state of AI technology, pose challenges and limitations. The authors used the latest version of LLMs at the time of conducting this study. However, AI models are advancing rapidly, and better models are being released in the computer science domain. Researchers in the applied research field should continue to be interested in the latest AI technologies and endeavor to apply them to solve problems in their field. Lastly, handling LLMs requires a large amount of computational resources. Due to the limited equipment and budget available, a medium-sized LLM was used in this study. In general, a larger model will have better performance. However, a developer should also consider memory and cost efficiency for practical use. The trade-off between performance and resource efficiency should be considered when developing an AI-driven KM tool. Moreover, researchers are investigating methods to maintain performance while making LLM lightweight and continue to explore on-device AI, which runs AI directly on local devices beyond the cloud server. Currently, a limited number of companies and professionals who have secured the necessary resources and equipment can use advanced AI technologies. But in the future, many companies and professionals, including developing countries, will be able to benefit from AI-driven KM tools as technology advances and development costs decrease.

With AI research and development, the fine-tuning process will be applicable to more advanced models, potentially enhancing the effectiveness and capabilities of the AI model in the complex engineering industry. In light of these ideas, more extensive and in-depth research is expected to develop advanced and efficient KM tools. In the future study, the authors will endeavor to develop an LLM-based in-house KM system that covers diverse subjects of technical specifications while limiting the leakage of internal data.

Author Contributions: Conceptualization, J.L., S.B. and W.J.; methodology, J.L. and S.B.; data curation, J.L. and S.B.; writing—original draft preparation, J.L., S.B. and W.J.; writing—review and editing, J.L., S.B. and W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the 2023 research fund of the KEPCO International Nuclear graduate School (KINGS), Republic of Korea and supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2020R1A2C1012739).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request because of confidentiality.

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

References

1. Moon, S.; Lee, G.; Chi, S. Automated System for Construction Specification Review Using Natural Language Processing. *Adv. Eng. Inform.* **2022**, *51*, 101495. [CrossRef]
2. Kivrak, S.; Arslan, G.; Dikmen, I.; Birgonul, M.T. Capturing Knowledge in Construction Projects: Knowledge Platform for Contractors. *J. Manag. Eng.* **2008**, *24*, 87–95. [CrossRef]
3. Bughin, J.; Chui, M.; Manyika, J. Capturing Business Value with Social Technologies. *McKinsey Q.* **2012**, *4*, 72–80.
4. Paul, C.; Patricia, C. Knowledge Management to Learning Organization Connection. *J. Manag. Eng.* **2007**, *23*, 122–130. [CrossRef]
5. Bigelow, B.F.; Perrenoud, A.J.; Rahman, M.; Saseendran, A. An Exploration of Age on Attraction and Retention of Managerial Workforce in the Electrical Construction Industry in the United States. *Int. J. Constr. Educ. Res.* **2021**, *17*, 3–17. [CrossRef]
6. Ashkenas, R. How to Preserve Institutional Knowledge. Available online: https://hbr.org/2013/03/how-to-preserve-institutional (accessed on 3 October 2023).
7. Lin, Y.-C.; Lee, H.-Y. Developing Project Communities of Practice-Based Knowledge Management System in Construction. *Autom. Constr.* **2012**, *22*, 422–432. [CrossRef]
8. Meese, N.; McMahon, C. Knowledge Sharing for Sustainable Development in Civil Engineering: A Systematic Review. *AI Soc.* **2012**, *27*, 437–449. [CrossRef]
9. Patricia, C.; Paul, C. Exploiting Knowledge Management: The Engineering and Construction Perspective. *J. Manag. Eng.* **2006**, *22*, 2–10. [CrossRef]
10. Amy, J.-W. Motivating Knowledge Sharing in Engineering and Construction Organizations: Power of Social Motivations. *J. Manag. Eng.* **2012**, *28*, 193–202. [CrossRef]
11. Salama, M.D.; El-Gohary, M.N. Semantic Text Classification for Supporting Automated Compliance Checking in Construction. *J. Comput. Civ. Eng.* **2016**, *30*, 4014106. [CrossRef]
12. Salama, A.D.; El-Gohary, M.N. Automated Compliance Checking of Construction Operation Plans Using a Deontology for the Construction Domain. *J. Comput. Civ. Eng.* **2013**, *27*, 681–698. [CrossRef]
13. Malsane, S.; Matthews, J.; Lockley, S.; Love, P.E.D.; Greenwood, D. Development of an Object Model for Automated Compliance Checking. *Autom. Constr.* **2015**, *49*, 51–58. [CrossRef]
14. Moon, S.; Lee, G.; Chi, S.; Oh, H. Automated Construction Specification Review with Named Entity Recognition Using Natural Language Processing. *J. Constr. Eng. Manag.* **2021**, *147*, 04020147. [CrossRef]
15. Belzner, L.; Gabor, T.; Wirsing, M. Large Language Model Assisted Software Engineering: Prospects, Challenges, and a Case Study. In Proceedings of the International Conference on Bridging the Gap between AI and Reality, Crete, Greece, 23–28 October 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 355–374.
16. Hassani, H.; Silva, E.S. The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data Cogn. Comput.* **2023**, *7*, 62. [CrossRef]
17. Gold, A.H.; Malhotra, A.; Segars, A.H. Knowledge Management: An Organizational Capabilities Perspective. *J. Manag. Inf. Syst.* **2001**, *18*, 185–214. [CrossRef]
18. Yacine, R. Ontology-Centered Knowledge Management Using Information Retrieval Techniques. *J. Comput. Civ. Eng.* **2006**, *20*, 261–270. [CrossRef]
19. Oppert, M.L.; O'Keeffe, V. The Future of the Ageing Workforce in Engineering: Relics or Resources? *Aust. J. Multi-Discip. Eng.* **2019**, *15*, 100–111. [CrossRef]
20. Wang, W.Y.C.; Pauleen, D.; Taskin, N. Enterprise Systems, Emerging Technologies, and the Data-Driven Knowledge Organisation. *Knowl. Manag. Res. Pract.* **2022**, *20*, 2039571. [CrossRef]
21. Chen, Y.; Luo, H.; Chen, J.; Guo, Y. Building Data-Driven Dynamic Capabilities to Arrest Knowledge Hiding: A Knowledge Management Perspective. *J. Bus. Res.* **2022**, *139*, 1138–1154. [CrossRef]
22. Kim, S.B. Impacts of Knowledge Management on the Organizationlal Success. *KSCE J. Civ. Eng.* **2014**, *18*, 1609–1617. [CrossRef]
23. Park, M.; Jang, Y.; Lee, H.-S.; Ahn, C.; Yoon, Y.-S. Application of Knowledge Management Technologies in Korean Small and Medium-Sized Construction Companies. *KSCE J. Civ. Eng.* **2013**, *17*, 22–32. [CrossRef]
24. Tan, H.C.; Carrillo, P.M.; Anumba, C.J. Case Study of Knowledge Management Implementation in a Medium-Sized Construction Sector Firm. *J. Manag. Eng.* **2012**, *28*, 338–347. [CrossRef]
25. Kale, S.; Karaman, E.A. A Diagnostic Model for Assessing the Knowledge Management Practices of Construction Firms. *KSCE J. Civ. Eng.* **2012**, *16*, 526–537. [CrossRef]
26. Hallowell, M.R. Safety-Knowledge Management in American Construction Organizations. *J. Manag. Eng.* **2012**, *28*, 203–211. [CrossRef]
27. Lauriola, I.; Lavelli, A.; Aiolli, F. An Introduction to Deep Learning in Natural Language Processing: Models, Techniques, and Tools. *Neurocomputing* **2022**, *470*, 443–456. [CrossRef]
28. Gunasekara, C.; Chalifour, N.; Triff, M. Question Answering Artificial Intelligence Chatbot on Military Dress Policy. Available online: https://cradpdf.drdc-rddc.gc.ca/PDFS/unc377/p813939_A1b.pdf (accessed on 27 January 2024).
29. 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. [CrossRef]

30. Kim, J.; Chung, S.; Moon, S.; Chi, S. Feasibility Study of a BERT-Based Question Answering Chatbot for Information Retrieval from Construction Specifications. In Proceedings of the 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Kuala Lumpur, Malaysia, 7–10 December 2022; pp. 970–974.
31. Teubner, T.; Flath, C.M.; Weinhardt, C.; van der Aalst, W.; Hinz, O. Welcome to the Era of ChatGPT et al.: The Prospects of Large Language Models. *Bus. Inf. Syst. Eng.* **2023**, *65*, 95–101. [CrossRef]
32. Liu, X.; Zheng, Y.; Du, Z.; Ding, M.; Qian, Y.; Yang, Z.; Tang, J. GPT Understands, Too. *AI Open* **2023**. [CrossRef]
33. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z. A Survey of Large Language Models. *arXiv* **2023**, arXiv:2303.18223.
34. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A. Language Models Are Few-Shot Learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
36. Zhuk, A. Navigating the Legal Landscape of AI Copyright: A Comparative Analysis of EU, US, and Chinese Approaches. *AI Ethics* **2023**, 1–8. [CrossRef]
37. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F. LLaMA: Open and Efficient Foundation Language Models. *arXiv* **2023**, arXiv:2302.13971.
38. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S. LLaMA 2: Open Foundation and Fine-Tuned Chat Models. *arXiv* **2023**, arXiv:2307.09288.
39. Xu, R.; Luo, F.; Zhang, Z.; Tan, C.; Chang, B.; Huang, S.; Huang, F. Raise a Child in Large Language Model: Towards Effective and Generalizable Fine-Tuning. *arXiv* **2021**, arXiv:2109.05687.
40. Lialin, V.; Deshpande, V.; Rumshisky, A. Scaling down to Scale up: A Guide to Parameter-Efficient Fine-Tuning. *arXiv* **2023**, arXiv:2303.15647.
41. Yuan, X.; Wang, T.; Wang, Y.-H.; Fine, E.; Abdelghani, R.; Lucas, P.; Sauzéon, H.; Oudeyer, P.-Y. Selecting Better Samples from Pre-Trained LLMs: A Case Study on Question Generation. *arXiv* **2022**, arXiv:2209.11000.
42. Dettmers, T.; Pagnoni, A.; Holtzman, A.; Zettlemoyer, L. QLoRA: Efficient Finetuning of Quantized LLMs. *arXiv* **2023**, arXiv:2305.14314.
43. OpenAI API Reference—Create Chat Completion. Available online: https://platform.openai.com/docs/api-reference/chat/create (accessed on 9 September 2023).
44. Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; De Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; Gelly, S. Parameter-Efficient Transfer Learning for NLP. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 2790–2799.
45. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. *arXiv* **2021**, arXiv:2106.09685.
46. Liu, H.; Tam, D.; Muqeeth, M.; Mohta, J.; Huang, T.; Bansal, M.; Raffel, C.A. Few-Shot Parameter-Efficient Fine-Tuning Is Better and Cheaper than in-Context Learning. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 1950–1965. [CrossRef]
47. OpenAI. GPT-4 Technical Report. *arXiv* **2023**, arXiv:2303.08774.

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.