

Article

A Comparison of Human Capabilities and Large Language Models for Knowledge Representation with Ontologies of Non-Destructive Testing in Bridge Engineering

Jan-Iwo Jäkel ^{1,*} , Eva Heinlein ¹ , Joy Sengupta ¹, Hongjo Kim ² and Katharina Klemt-Albert ¹

¹ Chair and Institute for Construction Management, Digital Engineering and Robotics in Construction, RWTH Aachen University, 52062 Aachen, Germany; eva.heinlein@rwth-aachen.de (E.H.); joy.sengupta@rwth-aachen.de (J.S.); klemt-albert@icom.rwth-aachen.de (K.K.-A.)

² Smart Infrastructure Laboratory, Department of Civil & Environmental Engineering, Yonsei University, Seoul 03722, Republic of Korea; hongjo@yonsei.ac.kr

* Correspondence: jaekel@icom.rwth-aachen.de

Abstract

Bridge structures are considered complex and significant. Accordingly, the knowledge of the engineering domain of bridge construction and related specialist areas is multidimensional and highly specific. Sometimes this knowledge is explicitly documented in standards, technical regulations, or information sheets. At other times, it resides implicitly in the expertise of the specialists involved. Ontologies are used to structure and formalize such domain knowledge, but creating them is resource-intensive and requires specialized expertise. Large language models (LLMs) offer one way to automate ontology creation through their natural language processing capabilities. This article examines LLMs' ability to generate ontologies in the specialized field of structural non-destructive testing (NDT) in bridge construction. Four different LLM-based approaches are employed. The results are compared with a previously created human-generated ontology and subsequently evaluated by external experts. Experts rate the human-developed SODIA ontology highest, with an average score of 3.44 out of 5 points. Only the ChatGPT 4.0-created ontology performed similarly well, with a score of 3.3 out of 5.00. All other LLM-based ontologies with ratings below 3.0 are of minor quality. These results underscore the potential and constraints of using LLMs to structure and formalize engineering domain knowledge into ontologies.

Keywords: large language models; artificial intelligence; ontology; knowledge representation; structural diagnostics



Academic Editor: Jaewook Jeong

Received: 9 January 2026

Revised: 13 March 2026

Accepted: 18 March 2026

Published: 1 April 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Bridge structures are central components of the transportation infrastructure, often designed for service lives of up to 100 years [1,2]. With the increasing traffic load in recent years, the structures have also come under increasing stress. As a result, there has been an increase in the occurrence of damages [3,4]. To better understand the causes and effects of such damage, additional diagnostic measures are deployed [5,6]. These diagnostic measures can be destructive, semi-destructive or non-destructive for the structure [7]. Non-destructive testing (NDT) methods have gained growing attention in both science and industry due to their minimal invasiveness [5,8]. Nevertheless, the NDT domain lacks comprehensive standardization and features a wide variety of highly specialized,

complex measures, each with different influencing parameters [9]. Because most NDT processes remain largely manual and resource-intensive, digitalization is hindered, impeding automated evaluation and interoperable data usage [8,10].

In addition to deploying digital tools and methods (e.g., Building Information Modeling (BIM)) to further digitize NDT processes, ontologies can play a central role by formalizing domain knowledge and enabling machine interpretability [11–13]. Creating ontologies is time-intensive and requires synthesizing both explicit and implicit knowledge. To address the complexity of developing ontologies, large language models (LLMs) offer a potential solution. LLMs can process large volumes of text and generate structured representations of domain-specific knowledge [14]. Yet, while LLMs have shown promise in other fields, their application in constructing ontologies for NDT diagnostics in bridge engineering remains underexplored.

In civil engineering, particularly in bridge construction and structural diagnostics, knowledge is highly complex, tightly interlinked, and subject-specific [9,15]. The formalization and structuring of this knowledge into ontologies relies on qualified expert engineers and is both labor- and time-intensive [16].

Despite the technological advances in [5], there is a problem in the scalability and standardization of knowledge representation in the specific domain. The current practice of knowledge formalization is characterized by fragmentation. The subject-specific knowledge about NDT processes is only provided in digitized PDF documents or is available to experienced engineers in unstructured form and as implicit expert knowledge [15,17].

The specific problem of this study can be divided into two aspects:

- Problem statement 1: The manual creation of ontologies by diagnostic experts is extremely time-consuming and costly due to the high technical complexity of nondestructive evaluation (NDE) in bridge construction. This hinders the comprehensive digitization of diagnostic processes toward machine-interpretable data.
- Problem statement 2: There is still a lack of reliable research on whether automated AI methods, such as LLMs, can fulfill the complex interrelationships and requirements for logical consistency and technical precision in safety-critical domains, such as bridge construction.

The aim of this article is to evaluate the efficiency of LLMs in the automated generation of domain-specific ontologies for NDE in bridge engineering. By benchmarking the LLM results against a validated, expert-developed reference ontology (SODIA), this study evaluates the current state of the art in automated knowledge formalization for a specialized engineering domain. The study focuses on the following research questions (RQ):

- RQ1: To what extent are LLMs able to correctly categorize and structure the multi-layered entities and influencing parameters of NDT in bridge construction and to logically define dependencies?
- RQ2: Can LLMs with the current performance formalize existing explicit knowledge in a subject-specific engineering domain better in an ontology automatically than an experienced specialist engineer can do manually?

2. Related Works

2.1. Ontologies in the Operation Phase of Bridge Structures

BIM has become increasingly prevalent during the planning and execution phases of infrastructure projects, including bridge construction. In parallel, numerous approaches have explored the use of BIM models for the operation and maintenance management of bridge structures, culminating in standardized BIM use cases [2]. However, due to limitations of the open data format Industry Foundation Classes (IFC) for capturing all

aspects of bridge operations, researchers have turned their attention toward alternative data integration and networking methods such as linked data. With the increasing attractiveness of the linked data approach, specific ontologies have been developed for different civil engineering domains, such as bridge construction, road construction, dam construction, and steel construction [18–20]. The ontologies and linked data approaches for bridge construction are recently gaining attention [17].

A foundational effort in this domain is the Bridge Topology Ontology (BROT), which provides a base ontology describing bridge components, spatial relationships, and structural properties [21,22]. Other examples in bridge maintenance include the Bridge Maintenance Ontology (BRMO) [23] and the Concrete Bridge Rehabilitation Project Management Ontology (CBRPMO) [24]. Collectively, these ontologies provide a fundamental representation of bridge components along with key aspects of operation and maintenance, such as preservation and maintenance management, structural damages and structural inspection tasks. While the BRMO and CBRPMO have all the content of the domain in one ontology and operate at a very strategic level due to its holistic nature, the BROT continues to use specific ontologies as extensions for a high level of detail specificity in the detailed areas of bridge maintenance. Additional extensions further extend the scope of BROT. In terms of materials, the BMAT ontology covers the various material types within bridge structures [21,25]. The Damage Topology Ontology (DOT) was developed by Hamdan et al. for the further description of damage to bridge structures and their further parameters and the resulting maintenance measures [26,27]. While these ontologies provide a robust framework for classifying components and damage, their high degree of specialization often results in isolated knowledge silos that struggle to map the complex interdependencies of diagnostic procedures. Furthermore, the extreme heterogeneity of practical applications proves too complex for rigid ontological structures, making it difficult to capture all project-specific details seamlessly. Consequently, these models lack the necessary flexibility to holistically integrate dynamic experiential knowledge and the contextual decision-making rationale required throughout a bridge lifecycle.

In addition to holistic ontologies in the field of operation and maintenance management of bridges, there are also ontologies specifically designed for a knowledge domain. Hence, structural diagnosis measures for bridge construction are described in the publications of Schuler et al. [15,28] while the SODIA ontology is proposed specifically for the formalization of NDT procedures [9]. These two complement each other, but also contain a limitation in their scope, due to the high heterogeneity of applications of diagnostic procedures, their dependence on each other and on the bridge structure.

Moreover, several ontologies target structural health monitoring (SHM) of bridge structures. For instance, Tsialiamanis et al. present a preliminary approach for modeling SHM systems, their functionalities, and associated data, linking them to bridge structures [29]. Li et al., 2021 propose with the Bridge Structure and Health Monitoring Ontology (BSHM) a similar approach [30]. This ontology comprises a Bridge Structure Module (BSM) and a Bridge Health Monitoring Module (BHMM). The BHMM focuses on sensor types, monitoring approaches, and measurement outcomes.

These ontologies focus primarily on the technical classification of sensor data and component metadata, but often neglect the link to the underlying design logic and the engineers' empirical knowledge. In addition, their rigid structure makes it difficult to efficiently integrate dynamic, unstructured information, such as that found in inspection reports or complex damage analyses. Many ontologies in the field of operation and maintenance management of bridge structures either focus on a holistic strategic approach or consider a specific domain. Nevertheless, the deconstruction phase is neglected in these. In this context of deconstruction, the BriDeCon ontology by Jäkel et al. integrates elements

from existing ontologies on bridge construction and operation to manage heterogeneous datasets for more efficient deconstruction planning [31].

Traditional ontologies in bridge engineering are inherently constrained by their rigid formalizations and isolated structures, which fail to accommodate the high heterogeneity of diagnostic procedures and the fragmented, unstructured knowledge found in technical documentation. Consequently, LLMs are essential to provide a flexible semantic layer capable of synthesizing these disparate data sources into a context-aware knowledge base, the possibilities of which are systematically investigated in this article.

2.2. Use of Large Language Models in the Construction Industry

In recent years, research interest in LLM technology and its use in the construction industry has increased [32]. Studies from Saka et al. and Ghimire et al. highlight various advantages of applying LLMs throughout the lifecycle of a civil structure, including improved knowledge management, automated data processing during planning and construction, and predictive operations [33,34]. However, these studies also identify key technical and socio-cultural challenges, such as hallucinations, a lack of interoperability, high costs, insufficient development and operational expertise, and low trust or acceptance [34].

The potential applications of LLMs in construction projects are diverse [32,33]. For instance, researchers have explored integrating LLMs with BIM models to automate construction site planning [35], scheduling [36], cost estimation [37], site reporting [38] and life cycle analyses [39]. Additionally, LLMs are being used in document and contract management [40–42], health and safety [43–45] or risk management [46,47]. Beyond these general applications, LLMs show promise in specialized sub-domains such as bridge construction, where they can be used for engineering design and structural calculations [48,49], as well as for optimizing operation, maintenance, and inspection processes [50–52]. Besides the possibilities of using LLM for further automation of business processes and processing of different data, the systems can also be used for the development of ontologies and knowledge graphs [53–55]. This approach is used in various industries, such as material engineering [56]. In the construction industry, LLMs thus offer the potential to structure and formalize complex, domain-specific engineering knowledge into a machine-interpretable knowledge graph [57].

Despite extensive research ontologies for bridge engineering and growing use of LLMs in various construction processes, there is a notable gap in applying LLMs specifically to automate ontology creation for NDT diagnostics in bridge structures. While domain-specific ontologies address areas such as damage assessment, maintenance, and structural health monitoring, few studies focus on systematically generating and evaluating NDT ontologies. Similarly, existing LLM-based solutions in construction have mainly addressed tasks like scheduling, cost estimation, or risk management, leaving the potential of LLMs for structured NDT domain knowledge largely unexamined. This shortfall highlights the need to investigate and compare the effectiveness of LLM-derived ontologies with manually developed ones in capturing the complexity and nuances of NDT in bridge engineering.

3. Methodology

This study follows a comparative research design to evaluate the performance of LLMs in formalizing engineering knowledge. Instead of a purely descriptive presentation of the work results, a systematic benchmarking procedure is applied, which is divided into the phases of preparation of the database, model preparation, ontology generation and evaluation (see Figure 1).

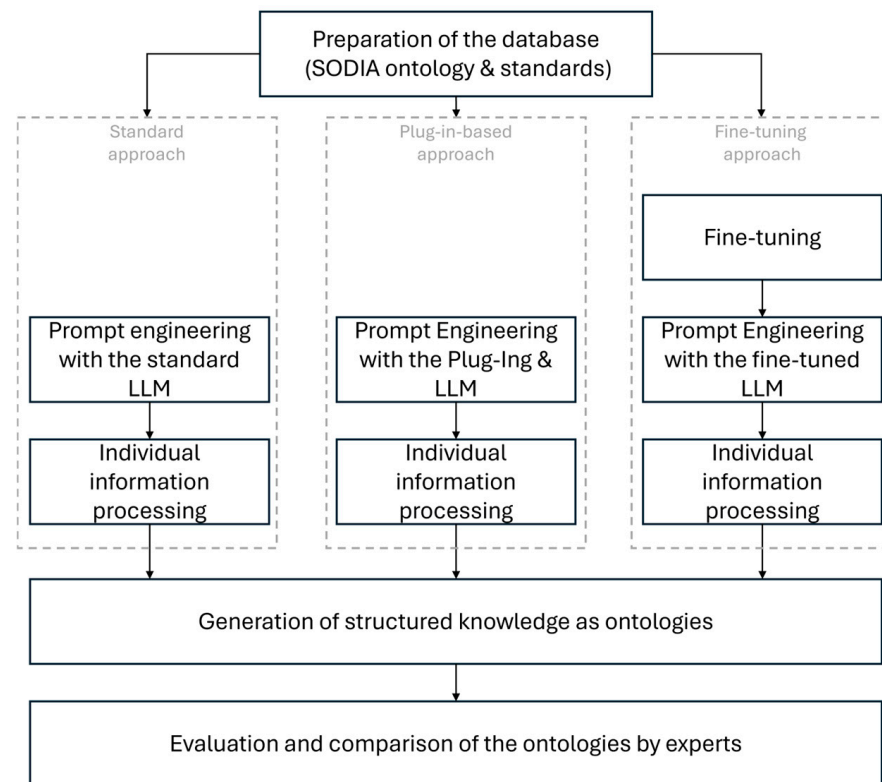


Figure 1. Procedure of the research.

The methodological basis of the study is a comparison between AI-generated ontologies and a validated reference instance. The primary data sources are a compilation of 16 reference documents, consisting of bridge construction standards, guidelines for non-destructive testing (NDT) and expert reports. These documents form the knowledge base from which the domain-specific concepts are extracted. The second data source is the SODIA ontology, which serves as a reference ontology. It forms the “ground truth” and was manually developed and practically validated based on the 16 reference documents in the articles [9,20]. This is available in the machine-readable Terse RDF Triple Language (TTL) format and acts as a qualitative basis against which the automatically generated structures are checked. To investigate the influence of different model architectures and optimization strategies, three configurations based on the GPT architecture from OpenAI are evaluated.

1. Standard approach: The standard LLM OpenAI GPT-3.5 and OpenAI GPT-4.0 are used.
2. Plug-in-based approach: Extension of OpenAI GPT-3.5 with the “Sider AI” [58] tool to test the added value of external tools in knowledge extraction.
3. Fine-tuning approach: Use of a specifically fine-tuned version of OpenAI GPT-3.5 [59,60].

The generation process is controlled by a standardized prompt engineering protocol. The chain-of-thought (CoT) method [61,62] is used here. This method enforces a step-by-step logical derivation of the concepts by the LLM. As a result, the complex causalities of engineering relationships in the NDT domain can be mapped more precisely by the LLMs.

After the generation process, the quality assurance of the generated ontologies is carried out by a multidimensional evaluation. A panel of experts consisting of three specialists in structural diagnostics was used for this purpose. Two are from science and one from industry. The choice of this composition ensures that both theoretical depth and practical applicability are included in the evaluation. The experts were sent the TTL files of

the ontologies for each individual approach for evaluation. They were also provided with guidelines for the qualitative evaluation of the results.

The evaluation is based on three criteria (i) Completeness, (ii) Practicability and (iii) Structure. These parameters are based on the evaluation approaches for ontologies of Gangemi et al. [63], Cristani and Cuel [64] in connection with Wilson et al. [65] and Bravo Contreras et al. [66]. They focus on evaluating the LLM performance for knowledge structuring and ontology generation. These three parameters collectively address the essential dimensions of domain coverage, operational utility, and logical integrity. Completeness ensures that all relevant aspects and correlations of NDT are represented through comprehensive data and object properties, thereby preventing critical knowledge gaps. Practicability assesses the proximity and industry relevance of formalized knowledge, ensuring the system provides actionable value for real-world operational projects. Finally, the parameter Structure evaluates the logical hierarchy and the meaningful definition of dependencies and restrictions between classes and parameters, which is vital for maintaining technical accuracy and formal consistency. Together, these three criteria provide a holistic and sufficient basis for determining whether a knowledge-based system can effectively meet the complex demands of bridge engineering.

The experts rate the LLM-generated ontologies individually on a scale from 1 (Very unsuitable) to 6 (Very suitable). In addition, an average value is formed for a ranking and the confidence interval for better comparability. Due to the number of only three experts, the Cohen kappa and Fleiss' kappa or a qualitative consensus process is not used. In this study, an average value and confidence intervals are utilized instead of Cohen's or Fleiss' kappa, because the small sample size of three experts does not meet the statistical requirements for stable reliability coefficients. In such limited cohorts, kappa values become disproportionately sensitive to minor disagreements, often producing misleading results that do not accurately reflect the true degree of consensus. Furthermore, this approach preserves the independence of individual expert judgments by avoiding the groupthink bias inherent in small-group qualitative consensus processes. Finally, the results are statistically analyzed and critically discussed regarding the suitability of LLMs for safety-critical engineering applications.

4. Introduction to the Research Base: SODIA Ontology

The SODIA ontology contains four main areas: (i) Inspection, (ii) Standardization, (iii) Formal Requirements and (iv) Influencing Factors, providing a structured framework for NDT in construction (see Figure 2). The SODIA ontology serves as the baseline for the study, providing a knowledge structure for NDT methods in bridge construction. It also serves as a reference ontology for later comparison between human-generated ontologies and various LLM-based ontologies in the field of NDE [9,17].

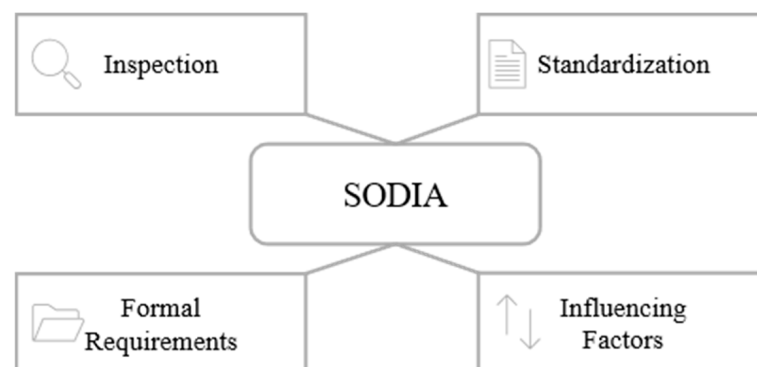


Figure 2. Structure of the SODIA Ontology.

The first main area, (i) Inspection, is the most comprehensive and is based primarily on the German guideline “Application of NDT methods in the construction industry” and the standard DIN 1076, which deals with the maintenance and inspection of infrastructural constructions such as bridges [67,68]. It includes not only inspection methods and tasks, but also related elements such as inspection tools, areas, and results. The concept is designed to be flexible and extensible, allowing the integration of specific NDT methods and related tasks in different contexts. The SODIA ontology can also accommodate different measurement principles, such as mechanical, electrical, and magnetic methods, as well as variations based on the material or geometry of the object under inspection.

The second main area, (ii) Standardization, defines the regulatory context and qualifications required for diagnostic projects. This includes the integration of standards such as DIN EN ISO 9712 [69], which specifies qualification requirements for personnel performing NDT, and technical regulations such as DIN 12504-2 [68] for the rebound hammer test. The SODIA ontology links inspection methods to the corresponding regulations and outlines the necessary qualifications for participants in diagnostic processes.

Next, the (iii) Formal Requirements segment of the SODIA ontology covers the systematic documentation of inspection results and the evaluation of structural conditions based on criteria such as durability, stability, and safety. The SODIA ontology includes classes for formal documentation of diagnostic procedures and findings, standardizing reporting and ensuring consistency across projects. Evaluation criteria are specified that allow infrastructure to be evaluated for structural integrity and safety, with the results integrated into the overall diagnostic workflow.

The last main area, (iv) Influencing Factors, considers constraints that may affect the applicability of inspection methods, such as chemical effects, structural conditions, and accessibility. These factors limit the selection of appropriate inspection methods; for example, the accessibility of a structure may limit the use of certain inspection methods.

5. LLM Approach

5.1. General Information About the Approaches

Two LLMs are used to generate the ontologies within the approach of the article. These are OpenAI’s LLM ChatGPT 3.5 and ChatGPT 4.0. The four LLM approaches—(a) ChatGPT 3.5 without plug-in, (b) ChatGPT 3.5 with plug-in, (c) ChatGPT 4.0 and (d) ChatGPT 3.5 fine-tuning—are derived based on the two AI models. Each approach generates a distinct LLM-based ontology in NDT for bridge structures. All approaches rely on a dual knowledge base. This dual base comprises (i) explicit knowledge drawn from norms, standards, technical regulations, and guidelines in structural diagnostics and bridge construction (see Table 1) and (ii) implicit knowledge gathered through six expert interviews conducted during the development and evaluation of the SODIA ontology [9,17]. As a result, the human-centered SODIA ontology and the four LLM-based ontologies share the same foundational knowledge, ensuring comparability at a later stage (see Chapter 6).

Table 1. Reference documents of the SODIA Ontology.

Pos.	Name (German Titles Have Been Translated into English)	Source
1	DIN 1076—Engineering structures in connection with roads—inspection and test	[67]
2.	Non-destructive testing methods for the determination of material parameters in reinforced and prestressed concrete construction	[70]
3.	Structural analysis	[71]

Table 1. *Cont.*

Pos.	Name (German Titles Have Been Translated into English)	Source
4.	Verification of modern non-destructive testing methods on a demolition structure	[72]
5.	Non-destructive testing—Qualification and certification of NDT personnel (ISO 9712:2021); German version: EN ISO 9712:2022	[69]
6.	Testing concrete in structures—Part 2: Non-destructive testing—Determination of rebound number; German version: DIN EN 12504-2:2021	[68]
7.	Measurement of the concrete cover and localization of the probaton	[73]
8.	Regulations in structural diagnostics	[74]
9.	Optimization of the remanence magnetism method for steel fracture detection in structures—theoretical further development	[75]
10.	Guidelines for the maintenance of engineering structures (RI-ERH-ING)	[76]
11.	Application of non-destructive testing methods in the construction industry (DBV Guideline)	[7]
12.	Code of practice for non-destructive concrete cover measurement and reinforcement location on reinforced and prestressed concrete components (Code of practice B02)	[77]
13.	Electrochemical potential measurements for the detection of reinforcement steel corrosion (Code of Practice B03)	[78]
14.	Ultrasonic methods for non-destructive testing in the construction industry (Code of practice B04)	[79]
15.	Leaflet on the radar method for non-destructive testing in the construction industry (Code of practice B10)	[80]
16.	Corrosion monitoring for reinforced and prestressed concrete structures (Code of practice B12)	[81]

The four LLM-based approaches are illustrated in Figure 3 and described in detail below:

- Approach (a)—ChatGPT 3.5 without plug-in: This is the standard version of ChatGPT 3.5 [82,83], employed in subsequent prompt engineering (see Section 5.3) without integrating any additional documents or applying fine-tuning.
- Approach (b)—ChatGPT 3.5 with plug-in: This approach extends (a) by employing the “Sider” plug-in [84] to integrate the explicit knowledge base (i.e., the relevant documents) directly into the LLM environment. Through this plug-in, ChatGPT 3.5 processes more domain-specific content on structural diagnostics and can refine its responses accordingly.
- Approach (c)—ChatGPT 4.0: This approach uses ChatGPT 4.0, which allows direct integration of text documents into its web-based interface. As with (b), the explicit knowledge base is incorporated to enhance the model’s domain-specific understanding of structural diagnostics.
- Approach (d)—ChatGPT 3.5 fine-tuning: This final approach utilizes a fine-tuned ChatGPT 3.5 on the dual knowledge base (combining both explicit and implicit knowledge). The fine-tuning steps, standardized prompt engineering procedure and subsequent generation of the LLM-based ontology are detailed in Sections 5.2 and 5.3.

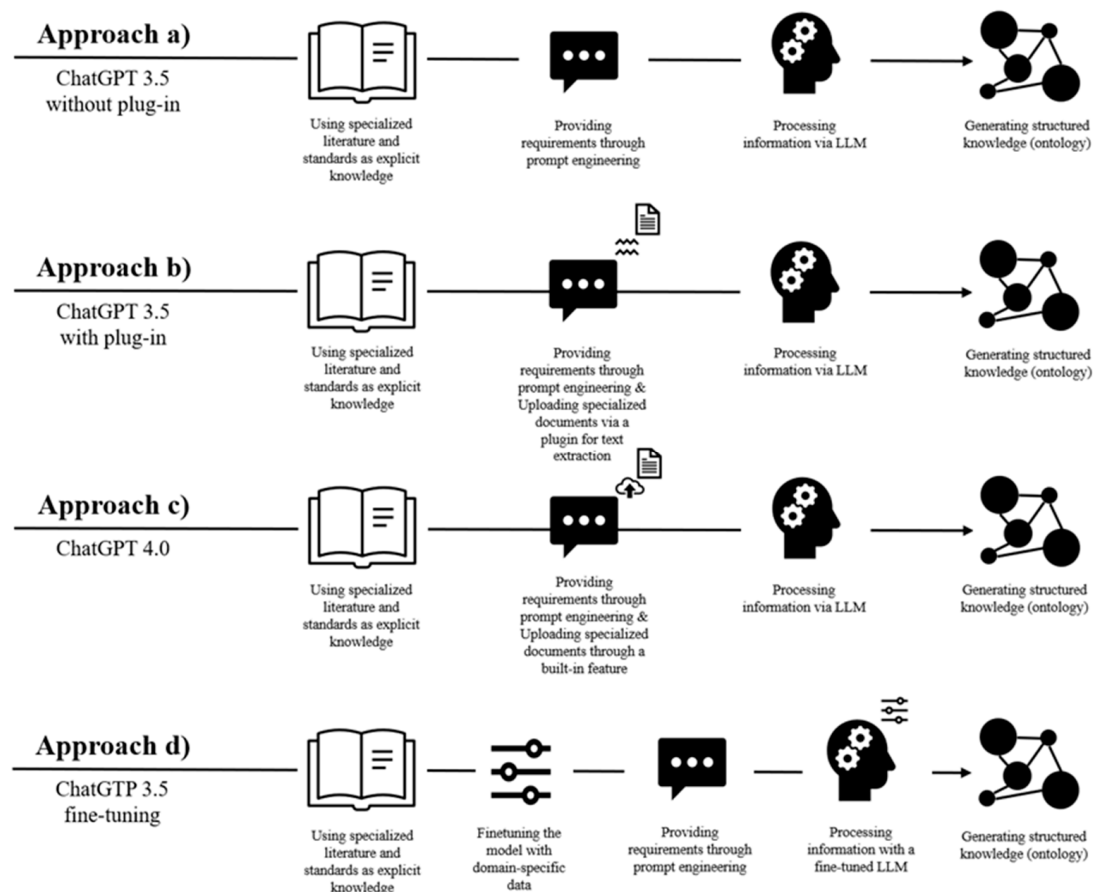


Figure 3. LLM-based approaches for the LLM-based ontology creation.

5.2. Fine-Tuning the LLM

While approaches (a) to (c) use the standard base models of ChatGPT 3.5 and ChatGPT 4.0, approach (d) employs an optimized model derived from a specialized fine-tuning process (see Figure 4). The fine-tuning consists of training and validation processes. The data set used for the fine-tuning comprises various sources of domain-specific knowledge (see Table 1) in the field of NDT methods for bridge structures. The aim of fine-tuning is to optimize the base GPT model into a specialized expert model. The OpenAI platform and its web-based graphical user interface are used to implement this process [85]. The data points have a standardized pattern consisting of a “comment” describing the domain under consideration and a message for the specified content. The content of the message is further differentiated into the parameters role, system, and content. An example from the dataset has the comment “NDT methods” and is structured as follows. The entire dataset was structured as a JSON file and used for fine-tuning the models.

Example of the question-and-answer format (Translated from German to English):

“_comment”: “Messmethode”

{“messages”: [{“role”: “system”, “content”: “ Experienced strucutruat diagnostician with a university degree and 10 years of professional experience. Has a very structured way of thinking.”}, {“role”: “user”, “content”: “ What measurement method does the rebound hammer use??”}, {“role”: “assistant”, “content”: “Penetration”}]}

Prior to integration into the LLM, the dataset is pre-processed. Approximately 100 reference data points are organized in a question-and-answer format. The data points got These data points are then supplied to the OpenAI platform as JSON and JSONL files structured with the roles System, User, and Assistant, and their corresponding content fields. The dataset is split into separate training (80%) and validation (20%) subsets, with

the validation set applied in a one-shot manner to evaluate the newly trained model. In addition to fundamental ontology knowledge for bridge maintenance, the dataset contains text from expert interviews and diagnostic sources, resulting in an uneven distribution of topics.

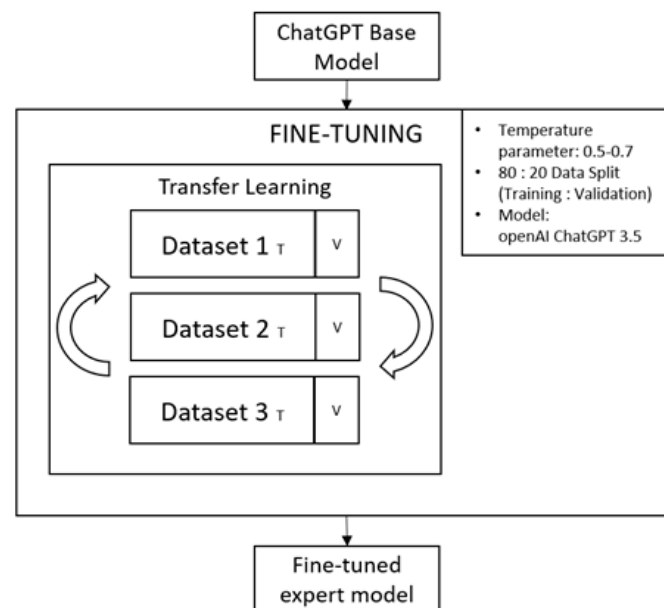


Figure 4. Fine-Tuning Approach with Transfer Learning.

To mitigate potential bias in the fine-tuned model, each topic is proportionally included in both training and validation subsets, reducing the risk of overfitting. Building on the conclusions of [86], the fine-tuning process also incorporates basic transfer-learning strategies such as training on similarly specialized domains and using pre-trained models. This study tested sequential fine-tuning (i.e., tuning the model on one dataset and then re-tuning with additional datasets), a technique referred to as continued pre-tuning in [87]. This experimentation yielded several model variants, including single-tuned, dual-tuned, and a final, iteratively refined model. The goal was to first establish a foundational ontology-based understanding of bridge maintenance, then incrementally incorporate advanced structural diagnostics data.

Because ChatGPT 3.5 makes its training parameters accessible, this study also experimented with varying the “temperature” parameter to balance the specificity and creativity of responses when compared to the standard ChatGPT 3.5 model. The OpenAI platform automatically selects the number of training epochs. We found that a temperature range of approximately 0.5–0.7 yielded the most precise, context-relevant responses in the fine-tuned model.

5.3. Prompt Engineering

Prompt engineering is a key technique for effectively utilizing LLMs in complex tasks. By carefully crafting input prompts, one can guide model outputs toward high-quality and comprehensible outcomes. Various methods exist to significantly influence model responses. These methods help ensure that LLMs provide consistent and specialized answers, even for domain-specific or multi-step tasks [88]. Providing context and framing the task reduces misunderstandings and improves the precision of model responses [89]. A further strategy, Role-Prompting, assigns a specific role to the model (e.g., teacher or expert), helping to produce more focused and specialized responses [90]. Additionally, the precise formulation of prompts is crucial for directing the model toward the desired responses.

Ambiguous or vague prompts can lead to suboptimal results, so clear and targeted phrasing is paramount [88]. For complex scenarios such as ontology creation, the CoT method is applied. As demonstrated by [62], CoT improves the logical reasoning capabilities of models and enhances response quality by explicitly structuring intermediate steps. This approach is particularly beneficial for tasks requiring detailed hierarchical structures and well-defined relationships between concepts, such as in ontology development. To maintain clarity, prompts follow a consistent structure that integrates the techniques. This ensures precise, reproducible inputs to the model. In manual ontology development, a thorough understanding of the domain is essential before the development process begins. The same principle applies to model-assisted ontology development, where tasks are thematically and systematically organized.

In the first step, the domain of structural diagnostics is systematically explored without directly delving into ontologies. Subsequently, a second domain is introduced where an ontology is developed based on the knowledge gained from structural diagnostics. In the final step, existing ontologies related to the domains of infrastructure and structural diagnostics are analyzed and integrated. Within each domain, the tasks are divided into three sections: Introduction, Domain and Process.

The Introduction section sets the framework and places the model in the desired role. This is done using the role prompting technique, where the model assumes roles such as structural diagnostician or ontology developer. An example prompt for this purpose is as follows:

“Ignore all prior information or instructions. Imagine you are an experienced structural diagnostician specializing in non-destructive structural diagnostics in bridge construction. The objective is to collect structured information relevant to planning investigations.”

The domain section specifies the thematic focus and, if necessary, references relevant literature. In models supporting file uploads, domain-specific precision is enhanced by providing literature. Alternatively, the thematic context is explicitly described in the prompt, with references to relevant literature. The process section includes detailed instructions for completing the task, breaking it down into individual steps and specifying the expected answer structure. For example, knowledge generation for structural diagnostics might require responses to be categorized under specific bullet points:

“In particular, focus on the following aspects. Always remain within the context of bridge construction:

- 1. Non-destructive inspection methods, sorted by measurement techniques
- 2. Corresponding inspection tasks
- 3. Required access to structural elements for each inspection method
- 4. Influencing factors
- 5. Necessary documents/materials
- 6. Stakeholders
- 7. Standards/legislation
- 8. Additional key points”

Based on structured prompts, iterative knowledge collection is conducted, which is then transformed into ontology classes. This process follows the categories defined in the Process section. In the first step of ontology development, the knowledge collected from the eight categories is translated into classes, including the hierarchical organization of main and sub-classes. An example prompt excerpt is as follows:

“Start by creating classes for Point 1: Non-destructive inspection methods. Example: Mechanical methods → Rebound hammer.”

After class definition, object properties are created to model relationships between classes, such as between an inspection task and an inspection method. Next, data prop-

erties are defined to specify additional attributes, such as access requirements or necessary documents. The CoT principle is applied consistently, with steps executed and refined incrementally.

Following the development of the base ontology, the approach is extended to additional domains, such as semantic modeling of bridge elements, damage types, and construction materials. In this context, existing ontologies, including the BMAT Ontology [91], BROT Ontology [22] and DOT Ontology [27] are referenced. This enables the adaptation and reuse of pre-existing classes and properties for the specific use case [92]. An example of the prompt and LLM output is shown in Table A1. The results of the ontologies generated in one iteration based on the four approaches are displayed in Table 2.

Table 2. Quantitative comparison of the resulting ontologies.

Approach	File Size of Chat [KB]	File Size of Ontology [KB]	Axioms [pcs.]	Logical Axioms [pcs.]	Classes [pcs.]	Object Properties [pcs.]	Data Properties [pcs.]
ChatGPT 3.5 without plug-in	100	40	457	193	106	27	29
ChatGPT 3.5 with plug-in	957	39	527	269	112	41	41
ChatGPT 4.0	66	26	444	259	135	35	23
ChatGPT 3.5 fine-tuning	26	15	226	127	73	19	7
SODIA	analog	498	3974	2367	253	88	109

Furthermore, the ontologies are presented visually in figures in the Appendix A (see Figures A1–A4). The SODIA ontology also serves as a baseline for development and for later comparison. During the generation process, the four LLM approaches revealed different challenges. ChatGPT 3.5 without plug-in had difficulties maintaining the connection to the eight predefined categories (see Section 5.3). As a result, not all relevant ontology elements were fully covered, and the naming of classes and properties had to be revised several times. Similar problems occurred when using ChatGPT 3.5 with plug-in. In addition, the language within the ontology sometimes alternated between German and English, complicating consistency. Using the plug-in, which extracts text from the PDF files and integrates it into the chat, occasionally resulted in longer processing times and extended chat interactions. ChatGPT 4.0 demonstrated better consistency in handling previously modeled knowledge. A significant benefit was the ability to upload documents without significantly extending the chat. This model generated the highest number of classes, while most properties were created with ChatGPT 3.5 with plug-in. The fine-tuned ChatGPT 3.5 model generated the smallest number of classes, properties and axioms, which is reflected in the smaller file size. However, even in this case, the connection to previously established knowledge structures were not consistently maintained. In addition, the naming of classes and properties often deviated from ontology conventions, such as the inclusion of “/” characters in names, which caused validation errors in the ontology.

6. Quality Assessment

Following the development and application of the LLM approaches, a quality check is implemented with the involvement of three independent experts with fundamental expertise in the field of structural diagnostics. Two scientific experts and one scientific expert are involved. All involved experts already supported in the article [9,17] for the acquisition of implicit knowledge for the development of the SODIA ontology and its evaluation. Therefore, the experts with reference to the subject area and the SODIA ontology

can qualitatively assess the evaluation of the ontology generated by the LLM. To be able to compare all five ontologies, quality criteria need to be defined. The experts involved must evaluate the ontologies according to these criteria. The three quality criteria parameters of completeness, practicability and structure are defined in Table 3.

Table 3. Evaluation parameters of the ontology assessment by the experts.

Pos.	Name of the Quality Criteria	Description of the Quality Criteria
1.	Completeness	The completeness of an ontology refers to the representation of all relevant aspects, correlations and the subject area of structural diagnostics in data properties and object properties.
2.	Practicability	The practicability of an ontology considers the proximity and relevance to the industry of formalized and represented knowledge for possible use in operational projects.
3.	Structure	The structure of an ontology focuses on the logic and meaningfulness of the hierarchical structure as well as defined dependencies and restrictions between individual classes, objects and parameters.

For each ontology, a rating is given for each quality criterion on a scale from very unsuitable (value: 1) to very suitable (value: 6) (see Table 4).

Table 4. Rating scale of the ontology assessment.

Points [Pt.]	Descriptions
1 Point [Pt.]	Very unsuitable
2 Points [Pt.]	Unsuitable
3 Points [Pt.]	Partly unsuitable
4 Points [Pt.]	Partly suitable
5 Points [Pt.]	Suitable
6 Points [Pt.]	Very suitable

The three quality criteria allow the ontologies under consideration to be optimally compared with each other in detail. Finally, an average value is calculated in order to identify the most suitable ontology based on the experts' opinions. Table 5 presents the evaluation of the individual experts per developed ontology and quality criterion as well as the result as an average value.

Table 5. Quantitative comparison of the results for the relevant ontologies of the procedure.

Expert-No. (Ex-No.)	Approach	Quality Criteria			Average Value [Pt.]
		Completeness [Pt.]	Practicability [Pt.]	Structure [Pt.]	
Ex-No. 1	ChatGPT 3.5 without plug-in	2	3	5	3.3
	ChatGPT 3.5 with plug-in	3	3	3	3.0
	ChatGPT 4.0	4	4	4	4.0
	ChatGPT 3.5 fine-tuning	2	2	2	2.0
	SODIA	5	4	5	4.7

Table 5. Cont.

Expert-No. (Ex-No.)	Approach	Quality Criteria			Average Value [Pt.]
		Completeness [Pt.]	Practicability [Pt.]	Structure [Pt.]	
Ex-No. 2	ChatGPT 3.5 without plug-in	2	1	1	1.3
	ChatGPT 3.5 with plug-in	1	1	1	1.0
	ChatGPT 4.0	3	3	3	3.0
	ChatGPT 3.5 fine-tuning	2	1	3	2.0
	SODIA	2	2	2	2.0
Ex-No. 3	ChatGPT 3.5 without plug-in	3	2	2	2.3
	ChatGPT 3.5 with plug-in	2	3	3	2.7
	ChatGPT 4.0	2	3	4	3.0
	ChatGPT 3.5 fine-tuning	2	3	3	2.7
	SODIA	4	4	3	3.7

Before the quantification process is completed, the individual ontologies are evaluated in the context of expert assessments:

- ChatGPT 3.5 without plug-in:
The ontology created uses many meaningful and correct terms. However, the LLM occasionally invents terms. The ontology is written in a consistent language. In addition, some of the object and data attributes are not appropriately named or are incorrect. Furthermore, the system generates too many object properties. The practicality of the ontology is considered average. The hierarchical structure of the ontology is appropriate, practical, and acceptable. However, some hierarchical classifications are very arbitrary and technically incomplete.
- ChatGPT 3.5 with plug-in:
Many meaningful and correct terms are used. However, the LLM occasionally invents terms and does not name them consistently. Synonyms or other spellings are often used. In addition, the language used is not consistent. A serious error is the switching between English and German when naming the individual classes. With regard to the properties, the data properties are selected as unsuitable and impractical. Furthermore, too many object properties are generated that are not suitable for the field of non-destructive testing. This results in rather average practicality for the engineer. The hierarchical structure is considered better than in the previous version, “ChatGPT 3.5 with plug-in.” In addition, the classes, data, and object properties are more deeply structured.
- ChatGPT 4.0:
The ontology created by ChatGPT 4.0 performs significantly better than that of its two predecessors, “ChatGPT 3.5 with plug-in” and “ChatGPT 3.5 without plug-in.” The completeness of the classes, objects, and data properties is considered comprehensive and effective. However, in direct comparison to the class and data properties, too many object properties are generated in this version. In addition, there is a lack of consistent naming. Synonyms and, in some cases, different spellings are used

selectively. Practicality is considered good, so an application that creates added value for the economy is considered feasible. According to the experts, the structure of the ontology is appropriate, but there is a lack of generic structure in some specific subject areas. In this context, the system could structure the ontology even more deeply.

- **ChatGPT 3.5 fine-tuning:**
The ontology generated by ChatGPT 3.5 fine-tuning is considered incomplete. It only includes a very small number of generic classes related to NDT and related fields. In addition, it contains too few object properties and only a few relevant data properties. Most data properties are not correctly defined. Furthermore, the ontology contains invented content that is not based on real technical content. The ontology is practically unusable and is considered critical. The structure of the ontology is very simple and global. There are too few links. Nevertheless, the existing structure is acceptable. The practicality is considered good, so that an application that creates added value for the economy is considered feasible. According to the experts, the structure of the ontology is appropriate, but there is a lack of generic structure in some specific subject areas. In this context, the system could structure the ontology even more deeply.
- **SODIA ontology:**
The SODIA ontology was not generated by an LLM, but is based on human thought and was developed in articles [9,17]. The ontology is classified as complete and of sufficient quality. All relevant classes related to the discipline of NDT and related topics are covered. The defined object and data properties are sufficient in number and consistently correctly named. The practicality is rated as good to very good, although this also depends on the end user's objectives. The structure is considered sound and comprehensive.
An examination of the individual scores indicates that no ontology received the highest score of 6 Pt. in any quality criterion. In addition, a score of 5 Pt. was only awarded three times. These points are all assigned to Ex-No.1. In the further evaluation of all experts, all ontologies are rated as moderately suitable (scoring 3 or 4 Pt.) or unsuitable (scoring 1 or 2 Pt.). For an overall comparison of the quality criteria in which the ontologies under consideration differ, the mean values for each quality criterion are calculated for each ontology (see Table 6).

Table 6. Summary of results and ranking of the approaches.

Approach	Completeness [Pt.]	Practicability [Pt.]	Structure [Pt.]	Average Value [Pt.]	Ranking
ChatGPT 3.5 without plug-in	2.33	2.00	2.67	2.33	3
ChatGPT 3.5 with plug-in	2.00	2.33	2.33	2.22	4
ChatGPT 4.0	3.00	3.33	3.67	3.33	2
ChatGPT 3.5 fine-tuning	2.00	2.00	2.67	2.22	4
SODIA	3.67	3.33	3.33	3.44	1

All generated ontologies are rated at most as partially suitable and generally as less suitable. The human-generated SODIA ontology received the best rating with an average score of 3.44 Pt. In this ontology, all three quality criteria are rated between 3 Pt. (partially unsuitable) and 4 Pt. (partially suitable). The completeness of the SODIA ontology received the highest average score of the three quality criteria (Completeness: 3.67 Pt.). Overall, it represents a mediocre result according to the experts' opinions. Nevertheless, the SODIA ontology is ranked first in a direct comparison. In second place is the ontology using the LLM ChatGPT 4.0 for development. This approach receives an overall score of 3.33 Pt.

This score is only marginally lower than that of the SODIA ontology. The structure of the ChatGPT 4.0 ontology is rated higher than the structure of the SODIA ontology with an average of 3.67 Pt., while the practicability of the two top-ranked ontologies receives the same average score (Practicability: 3.33 Pt.), the completeness is regarded as lower (Completeness: 3.00 Pt.). With an average score of 2.33 points, the ontology of the ChatGPT 3.5 approach without plug-in was ranked third. All experts rated the generated ontology as partially unsuitable in the quality criteria of structure and completeness (Structure: 2.67 Pt.; Completeness: 2.33 Pt.) and unsuitable in its practicability (Practicability: 2.00 Pt.).

Based on the ranking based on the average points of the experts in the three quality criteria, it can be concluded that humans can currently generate better ontologies than intelligent AI models. This also results in the fact that human capabilities can better process, formalize, structure and contextualize existing explicit and implicit expert knowledge, in this context the waking knowledge of structural diagnostics, compared to AI models. Nevertheless, all generated ontologies, whether human or artificially generated, are not optimal and are evaluated as partially inappropriate or even inappropriate. Thus, the formalization of engineering knowledge and the development of a domain-specific ontology can still be characterized as a complex and challenging process.

In addition to evaluating the quality of all ontologies created based on expert evaluation, the following general errors can be derived for LLM-based ontologies:

- All approaches of ChatGPT 3.5 are accompanied by hallucinations of the system. The LLM invents new parameters that do not exist for the definition of classes and object attributes, such as new methods and tools in the field of structural diagnostics or non-existent standards.
- All approaches of ChatGPT 3.5 automatically generate more data properties than object properties when creating ontologies.
- All approaches of ChatGPT 3.5 use synonyms and different spellings when naming classes, object properties, and data properties.
- All approaches of ChatGPT 3.5 structure ontologies generically and only up to a second level. All ontologies lack further depth in the definition of dependencies.
- The ontology of ChatGPT 4 most closely resembles the human-generated ontology, both in terms of completeness and practicality. Only in terms of structuring is the human-generated ontology even more specific and meaningfully linked.

After a general comparison with the average, the 95% confidence interval is used for further analysis. The average values from Table 6 are used for the database. As the sample size is limited to $n = 3$ experts, the t-distribution is used. With a confidence level of 95% and 2 degrees of freedom, the critical t-value is 4.303. In addition, the standard deviation, the standard error and the margin of error including the 95% CI are calculated and displayed in Table 7.

The analysis of the expert evaluations highlights a variance between the raters, which leads to wide confidence intervals, particularly for the SODIA approach. This dispersion results primarily from the fact that Expert 2 consistently rated all approaches significantly lower than Experts 1 and 3. Regarding reliability, ChatGPT 3.5 fine-tuning shows the narrowest confidence interval ([1.23, 3.23]), suggesting the most consistent agreement among the experts for this specific method, despite its lower overall ranking. Among the top performers, although SODIA achieves the highest mean value of 3.47, its 95% confidence interval overlaps significantly with that of ChatGPT 4.0. This indicates that with a sample size of only three experts, the statistical difference between the two leading approaches cannot yet be regarded as definitive at a 95% confidence level.

Table 7. Calculation of the 95% confidence interval.

Approach	Standard Deviation (s)	Standard Error (SE)	Margin of Error	95% Confidence Interval (CI)
ChatGPT 3.5 without plug-in	1.00	0.58	± 2.48	[−0.18, 4.78]
ChatGPT 3.5 with plug-in	1.08	0.62	± 2.68	[−0.45, 4.91]
ChatGPT 4.0	0.58	0.33	± 1.44	[1.89, 4.77]
ChatGPT 3.5 fine-tuning	0.40	0.23	± 1.00	[1.23, 3.23]
SODIA	1.37	0.79	± 3.40	[0.07, 6.87]

7. Discussion

The study demonstrates the potential for automating the formalization and structuring of complex engineering knowledge using LLMs. This facilitates the rapid development of domain-specific ontologies and enhances data interoperability. In addressing RQ1, the findings indicate that while LLMs can effectively categorize high-level entities, they currently struggle to match the granularity and logical consistency of the expert-generated SODIA ontology. This discrepancy suggests that while LLMs offer efficiency gains in time and resource usage, they exhibit a limited capacity to mirror the technical depth achieved by human experts, likely due to “hallucinations” or a lack of deep semantic grounding in safety-critical diagnostic interdependencies. Regarding RQ2, the direct comparison confirms that current models cannot yet autonomously formalize subject-specific knowledge with the same precision as a specialist engineer. However, given the rapid advances in model parameter sizes and reasoning capabilities, it is plausible that LLMs will, in the medium term, more effectively process complex engineering relationships to produce more robust and nuanced ontologies.

For practitioners, these results imply that LLMs should currently be viewed as intelligent assistants rather than fully autonomous agents. In the context of Problem Statement 1, LLMs can be realistically integrated into the initial drafting stages of the ontology engineering lifecycle, serving as generalized assistants within digital BIM processes to transform unstructured legacy data into structured knowledge drafts. This hybrid “human-in-the-loop” workflow allows for the rapid generation of basic and generalized ontologies, which can then be refined and validated by experts to ensure technical accuracy and practical relevance for decision-making. Furthermore, the results fulfill the second problem statement, demonstrating that LLMs can generate basic NDT ontologies. Despite this, structures developed by humans, such as the SODIA ontology, remain superior in terms of completeness and practicality. While the tested GPT models showed potential, they generally failed to move beyond a generic level, lacking the necessary depth for complex, subject-specific engineering relations.

A critical limitation remains in the presence of “hallucinations,” where the systems invent non-existent NDT methods, standards, and inconsistent naming conventions. Statistical analysis confirms that even the top-performing AI approaches struggle to fully represent the intricate complexity and hierarchical requirements of structural diagnostics. Ultimately, the formalization of engineering knowledge remains a challenging process that currently requires human expert intervention to ensure technical accuracy and depth.

The scope of this study was primarily restricted to two proprietary models, ChatGPT 3.5 and 4.0, which limits the generalizability of the findings across the diverse landscape of LLMs. To develop a more holistic understanding, future research must benchmark these

results against a wider variety of architectures, including open-source models such as Meta LLaMA, Google BERT, Falcon-2 and Claude. These LLMs, particularly those with larger parameter sizes or specialized engineering pre-training, may offer different structural biases and superior preservation of complex conceptual relationships. Furthermore, the reliance on a relatively narrow knowledge base within a total of 16 reference documents and three experts for the validation remains a limitation. The manual design of prompts also introduces potential inconsistencies. Future work should transition toward automated prompt engineering and advanced reasoning frameworks to minimize bias and enhance logical transparency.

To achieve a robust validation, the current evaluation by a small expert cohort must be expanded. Future studies should employ larger expert panels, interdisciplinary workshops, or large-scale quantitative surveys to ensure statistical validity and practical applicability across the industry. Additionally, the integration of multi-modal datasets incorporating not only text, but also structural diagrams, NDT sensor logs, and visual inspection data is essential to overcome the limitations of purely text-based knowledge retrieval. Expanding this research into broader domains, such as damage management, structural recalculations, and maintenance planning, will be critical to fully project these AI capabilities onto the BIM method. These efforts should focus on developing hybrid workflows with a collaboration between human engineers and LLMs. In these workflows, LLMs act as scalable semantic layers, bridging the gap between fragmented legacy data and a fully interoperable, machine-readable management ecosystem.

8. Conclusions

The study confirms that, whilst current LLMs are able to successfully identify high-level entities and fundamental dependencies, their current capabilities are reaching their limits in the context of knowledge engineering, and they struggle to capture the complex, multi-layered interrelationships of formalized expert knowledge within the engineering discipline of structural diagnostics. Consequently, these models do not achieve the detailed precision and logical consistency that are characteristic of structures developed by experts. Therefore, expert support is required to ensure technical accuracy.

A direct comparative analysis shows that automated approaches using LLMs do not match the quality of manually generated ontologies. Although LLMs significantly reduce the ‘knowledge engineering bottleneck’ by cutting time and resource costs, the human-created base ontology remains superior in terms of technical completeness and practical applicability. Consequently, LLMs are currently best deployed as ‘intelligent semantic assistants’ within a hybrid, human-in-the-loop workflow, rather than as fully autonomous agents for formalizing subject-specific engineering knowledge. With the rapid advancement of language model capabilities in the future, the quality of ontology development outcomes may improve. Nevertheless, the specialist engineer should always have the final say on the ontologies before they are released and put into practice.

Author Contributions: J.-I.J.: Design, Methodology, Acquisition and interpretation of data, Drafting the article, Funding acquisition. E.H.: Acquisition and interpretation of data, Drafting the article, Reviewing the Article. J.S.: Development of the algorithm, Analysis of data, Drafting the article, Reviewing the Article. H.K.: Reviewing the Article, Final Approval, Supervision, K.K.-A.: Final Approval, Project Administration, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Federal Ministry for Digital and Transport (BMDV) under Grant 19FS2021A in the funding program mFUND in the project mdBIM+.

Table A1. Example of a prompt and output of the LLM (The main text was translated from German into English; the content of the ontology was not translated to represent the original).

Prompt of the User:

User:

Now classify all listed test procedures as respective subclasses.

Answer:

ChatGPT:

Here is the extension of the ontology to sort all listed test procedures as respective subclasses:

Copy code

@prefix: <<http://example.org/ontology#>>.

Continuation of ontology with specific test procedures

:Georadar rdf:type owl:Class;

rdfs:subClassOf:ElektrischeUndElektromagnetischeVerfahren.

:Wirbelstromprüfung rdf:type owl:Class;

rdfs:subClassOf:ElektrischeUndElektromagnetischeVerfahren.

:Betonradar rdf:type owl:Class;

rdfs:subClassOf:ElektrischeUndElektromagnetischeVerfahren.

:MagnetischeFlussdichte rdf:type owl:Class;

rdfs:subClassOf:MagnetischeVerfahren.

:Metalldetektion rdf:type owl:Class;

rdfs:subClassOf:MagnetischeVerfahren.

:Karbonatisierungstests rdf:type owl:Class;

rdfs:subClassOf:ChemischeUndElektrochemischeVerfahren.

:ChloridgehaltMessung rdf:type owl:Class;

rdfs:subClassOf:ChemischeUndElektrochemischeVerfahren.

:Potenzialfeldmessung rdf:type owl:Class;

rdfs:subClassOf:ChemischeUndElektrochemischeVerfahren.

:Thermografie rdf:type owl:Class;

rdfs:subClassOf:AndereVerfahren.

:Endoskopie rdf:type owl:Class;

rdfs:subClassOf:AndereVerfahren.

This supplement assigns the remaining test methods from the previously created list to the corresponding subclasses, such as ground-penetrating radar, eddy current testing, concrete radar, magnetic flux density, metal detection, carbonation tests, chloride content measurement, potential field measurement, thermography, and endoscopy. Each of these methods is assigned to a higher-level category based on its method or area of application.

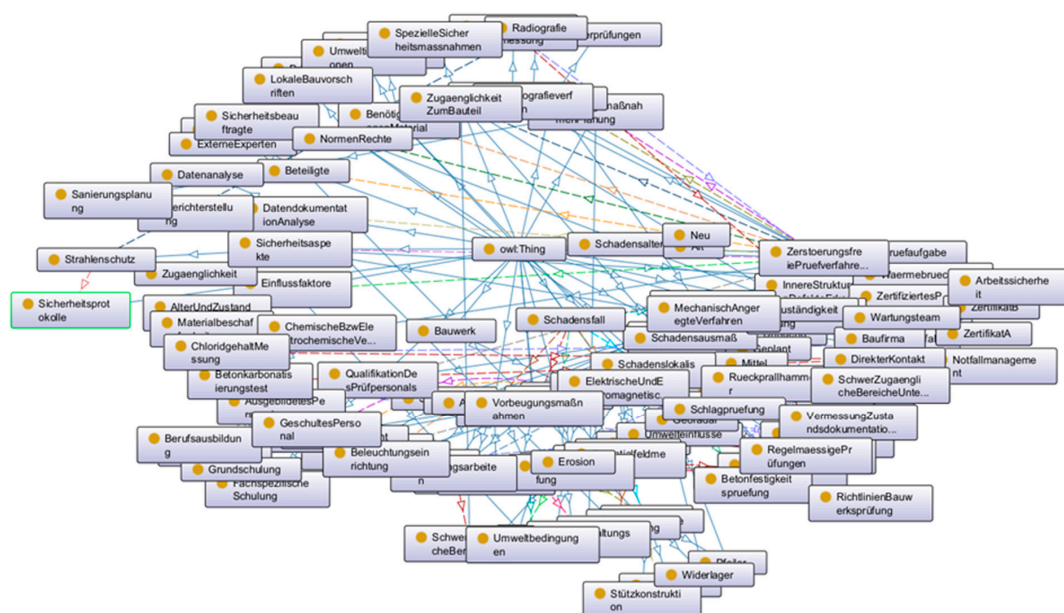


Figure A3. Visualization of the ontology from ChatGPT 4.0.

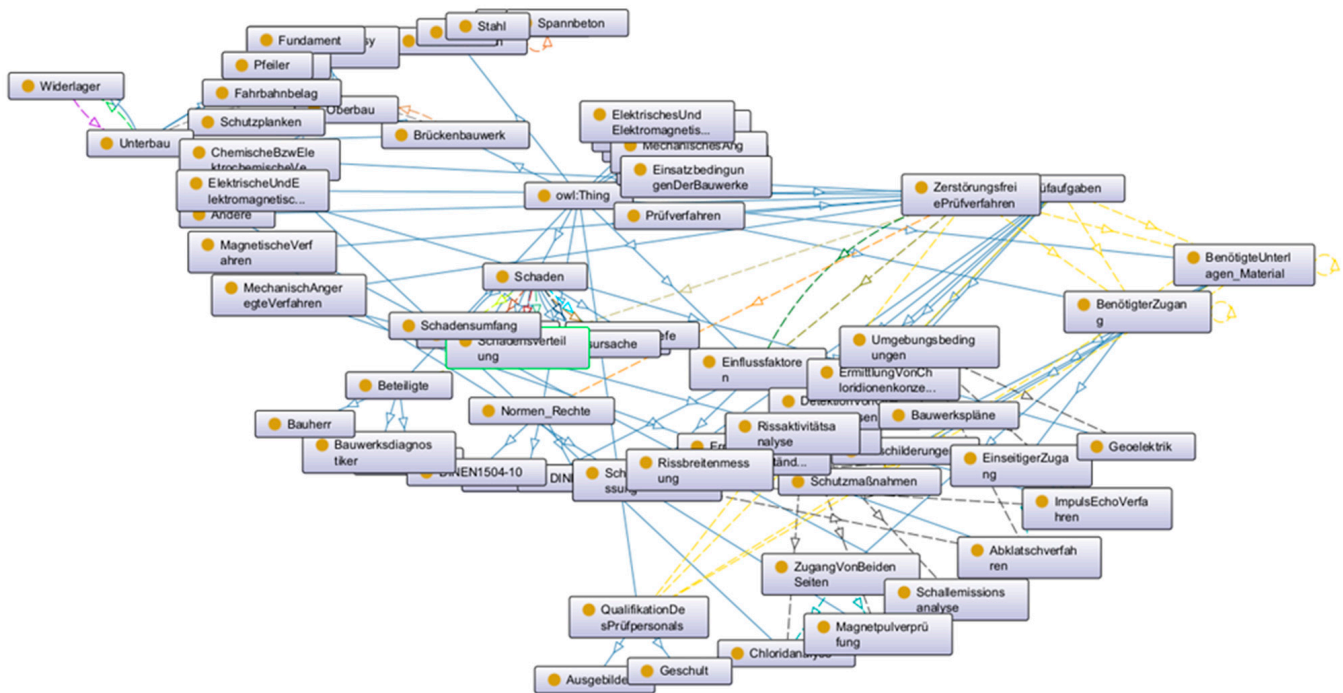


Figure A4. Visualization of the ontology from ChatGPT 3.5 fine-tuning.

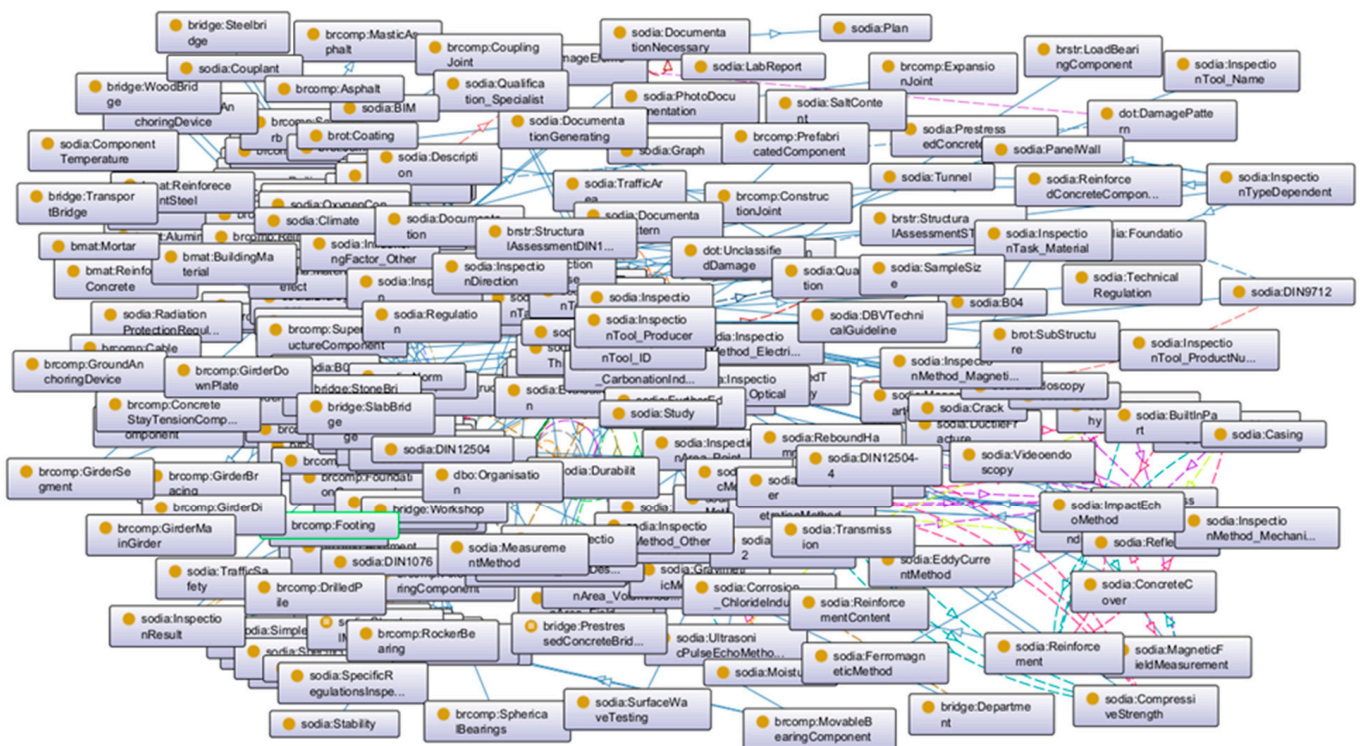


Figure A5. Visualization of the SODIA ontology.

References

1. Hartung, R.; Schönbach, R.; Liepe, D.; Klemt-Albert, K. Automated Parametric Modeling to Enhance a data-based Maintenance Process for Infrastructure Buildings. In *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC), Kitakyushu, Japan, 27–29 October 2020*; Osumi, H., Ed.; International Association for Automation and Robotics in Construction (IAARC): Oulu, Finland, 2020.
2. Jäkel, J.-I.; Klemt-Albert, K. BIM-models of bridges in the operational phase: Use cases, phase model and reference architecture. *ce/papers* **2023**, *6*, 701–710. [[CrossRef](#)]

3. Hartung, R.; Naraniecki, H.; Klemt-Albert, K.; Marx, S. Konzept zur BIM-basierten Instandhaltung von Ingenieurbauwerken mit Monitoringsystemen. *Bautechnik* **2020**, *97*, 826–835. [[CrossRef](#)]
4. Naraniecki, H.; Hartung, R.; Marx, S.; Klemt-Albert, K. Zustandsprognose von Ingenieurbauwerken auf Basis von digitalen Zwillingen und Bestandsdaten. *Bautechnik* **2022**, *99*, 173–181. [[CrossRef](#)]
5. Niederleithinger, E. NDE 4.0 in Civil Engineering. In *Handbook of Nondestructive Evaluation 4.0*; Meyendorf, N., Ida, N., Singh, R., Vrana, J., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 937–949, ISBN 978-3-030-73205-9.
6. Schickert, M.; Artus, M.; Koch, C. Integration and Visualization of NDE Data in Digital Building Models—A conceptual view. *E-J. Nondestruct. Test.* **2022**, *27*. [[CrossRef](#)] [[PubMed](#)]
7. German Concrete and Construction Engineering Association. *DBV-Merkblatt “Anwendung Zerstörungsfreier Prüfverfahren im Bauwesen” (Engl. Application of Non-Destructive Testing Methods in the Construction Industry’)*; Deutscher Beton- und Bautechnik-Verein E.V.: Berlin, Germany, 2014.
8. Arndt, R.; Voigt, C.; Fritsch, C.; Eric, E.; Burkhardt, J. Digital Building Diagnostics. *E-J. Nondestruct. Test.* **2022**, *27*. [[CrossRef](#)]
9. Jäkel, J.-I.; Heinlein, E.; Klemt-Albert, K. Formalisation of Knowledge About Non-Destructive Testing Methods For Bridge Structures Based on The Development of The Sodja Ontology. In *Proceedings of the Creative Construction Conference 2024, Praha, Czech Republic, 29 June–2 July 2024*; Budapest University of Technology and Economics: Budapest, Hungary, 2024.
10. Voigt, C.; Fritsch, C.; Hackel, T. Digitalisierung der Bauwerksdiagnostik zur realitätsnahen Bewertung von Ingenieurbauwerken. In *2. Fachkongress Digitale Transformation der Verkehrsinfrastruktur: Fachtagung über Planung, Bau, Betrieb, Unterhalt, Rückbau von Brücken, Tunneln, Schienen, Straßen, Wasserwegen Digital*; 1. Auflage; Krieger, J., Ed.; Expert Verlag Ein Imprint von Narr Francke Attempto Verlag: Tübingen, Germany, 2023; ISBN 978-3-8169-3554-4.
11. Morgenstern, H.; Raupach, M. A Novel Approach for Maintenance and Repair of Reinforced Concrete Using Building Information Modeling with Integrated Machine-Readable Diagnosis Data. *Constr. Mater.* **2022**, *2*, 314–327. [[CrossRef](#)]
12. Morgenstern, H.; Raupach, M. Predictive BIM with Integrated Bayesian Inference of Deterioration Models as a Four-Dimensional Decision Support Tool. *CivilEng* **2023**, *4*, 185–203. [[CrossRef](#)]
13. D’Amico, F.; Bianchini Ciampoli, L.; Di Benedetto, A.; Bertolini, L.; Napolitano, A. Integrating Non-Destructive Surveys into a Preliminary BIM-Oriented Digital Model for Possible Future Application in Road Pavements Management. *Infrastructures* **2022**, *7*, 10. [[CrossRef](#)]
14. Lo, A.; Jiang, A.Q.; Li, W.; Jamnik, M. End-to-End Ontology Learning with Large Language Models. *arXiv* **2024**, arXiv:2410.23584.
15. Schuler, P.-C.; Mirboland, M.; Voigt, C.; Koch, C. Using Linked Data Containers for BIM based structural inspection workflows. In *Advances in Information Technology in Civil and Building Engineering—Proceedings of the 20th Conference of the International Society for Computing in Civil and Building Engineering 2024—Volume 3: Construction Management, Montreal, Canada, 25 August 2024–28 August 2025*; Francis, A., Miresco, E., Melhado, S., Eds.; Springer Nature: Cham, Switzerland, 2024.
16. Funk, M.; Hosemann, S.; Jung, J.C.; Lutz, C. Towards Ontology Construction with Language Models. *arXiv* **2023**, arXiv:2309.09898. [[CrossRef](#)]
17. Jäkel, J.-I.; Heinlein, E.; Morgenstern, H.; Kim, H.; Klemt-Albert, K. System- and Data-integrated linking of digital 3D models of existing bridge structures with Knowledge Graphs of non-destructive diagnostic methods. *J. Inf. Technol. Constr.* **2025**, *30*, 603–630. [[CrossRef](#)]
18. Pauwels, P.; Krijnen, T.; Terkaj, W.; Beetz, J. Enhancing the ifcOWL ontology with an alternative representation for geometric data. *Autom. Constr.* **2017**, *80*, 77–94. [[CrossRef](#)]
19. Pauwels, P.; Costin, A.; Rasmussen, M.H. Knowledge Graphs and Linked Data for the Built Environment. In *Industry 4.0 for the Built Environment*; Bolpagni, M., Gavina, R., Ribeiro, D., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 157–183, ISBN 978-3-030-82429-7.
20. Pauwels, P.; Terkaj, W. EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Autom. Constr.* **2016**, *63*, 100–133. [[CrossRef](#)]
21. Hamdan, A.-H. *Ein Ontologiebasiertes Verfahren zur Automatisierten Bewertung von Bauwerksschäden in Einer Digitalen Datenumgebung*; 1. Auflage; Institut für Bauinformatik, Fakultät Bauingenieurwesen, TU Dresden: Dresden, Germany, 2023; ISBN 978-3-86780-728-9.
22. Hamdan, A.-H.; Kozak, T. Bridge Topology Ontology. Available online: <https://alhakam.github.io/brot/> (accessed on 3 January 2025).
23. Ren, G.; Ding, R.; Li, H. Building an ontological knowledgebase for bridge maintenance. *Adv. Eng. Softw.* **2019**, *130*, 24–40. [[CrossRef](#)]
24. Wu, C.; Wu, P.; Wang, J.; Jiang, R.; Chen, M.; Wang, X. Ontological knowledge base for concrete bridge rehabilitation project management. *Autom. Constr.* **2021**, *121*, 103428. [[CrossRef](#)]
25. Hamdan, A.-H.; Taraben, J.; Helmrich, M.; Mansperger, T.; Morgenthal, G.; Scherer, R.J. A semantic modeling approach for the automated detection and interpretation of structural damage. *Autom. Constr.* **2021**, *128*, 103739. [[CrossRef](#)]
26. Hamdan, A.; Scherer, R.J. A generic model for the digitalization of structural damage. In *Life Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision*; Caspeele, R., Taerwe, L., Frangopol, D.M., Eds.; CRC Press: London, UK, 2019; ISBN 9781315228914.

27. Hamdan, A.-H.; Bonduel, M. Damage Topology Ontology. Available online: <https://alhakam.github.io/dot/> (accessed on 3 January 2025).
28. Schuler, P.-C.; Stark, Y.; Mirboland, M. (Eds.) *Informationsintegration in der Betriebsphase von Bauwerken: Eine Diagnostikspezifische Ontologie im ICDD*; Technische Universität Hamburg, Institut für Digitales und Autonomes Bauen: Hamburg, Germany, 2024.
29. Tsialiamanis, G.P.; Wagg, D.J.; Antoniadou, I.; Worden, K. An Ontological Approach to Structural Health Monitoring. In *Topics in Modal Analysis & Testing*; Dilworth, B., Mains, M., Eds.; Springer International Publishing: Cham, Switzerland, 2021; Volume 8, pp. 51–59, ISBN 978-3-030-47716-5.
30. Li, R.; Mo, T.; Yang, J.; Jiang, S.; Li, T.; Liu, Y. Ontologies-Based Domain Knowledge Modeling and Heterogeneous Sensor Data Integration for Bridge Health Monitoring Systems. *IEEE Trans. Ind. Inf.* **2021**, *17*, 321–332. [[CrossRef](#)]
31. Jäkel, J.-I.; Heinlein, E.; Mackenbach, S.; Klemm-Albert, K. BIM-based deconstruction of bridge structures using digital 3D-models and information containers. In *Proceedings of the International Conference of Smart and Sustainable Built Environment (SASBE 2024)*, Auckland, New Zealand, 7–9 November 2024.
32. Taiwo, R.; Bello, I.T.; Abdulai, S.F.; Yussif, A.-M.; Salami, B.A.; Saka, A.; Zayed, T. Generative AI in the Construction Industry: A State-of-the-art Analysis. *arXiv* **2024**, arXiv:2402.09939.
33. 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.* **2024**, *17*, 100300. [[CrossRef](#)]
34. Ghimire, P.; Kim, K.; Acharya, M. Opportunities and Challenges of Generative AI in Construction Industry: Focusing on Adoption of Text-Based Models. *Buildings* **2024**, *14*, 220. [[CrossRef](#)]
35. Zhen, Y.; Bi, S.; Tang, S.; Lu, X.-t.; Pan, W.-q.; Shi, H.-p.; Chen, Z.-r.; Fang, Y.-s.; Wang, X.-m. LLM-Project: Automated Engineering Task Planning via Generative AI and WBS Integration. In *2024 IEEE 14th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Copenhagen, Denmark, 16–19 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 605–610, ISBN 979-8-3315-0605-6.
36. 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. [[CrossRef](#)]
37. Parsafard, P.; Elezaj, O.; Ekundayo, D.; Vakaj, E.; Parmar, M.; Ahmad Wani, M. Automation in Construction Cost Budgeting using Generative Artificial Intelligence. In *Proceedings of the 14th International Conference on Industrial Engineering and Operations Management*, Dubai, United Arab Emirates, 14–16 November 2023; IEOM Society International: Southfield, MI, USA, 2024; ISBN 979-8-3507-1734-1.
38. Pu, H.; Yang, X.; Li, J.; Guo, R. AutoRepo: A general framework for multimodal LLM-based automated construction reporting. *Expert Syst. Appl.* **2024**, *255*, 124601. [[CrossRef](#)]
39. Chen, L.; Silvennoinen, H.; de Wolf, C.; Hall, D.; van Mele, T.; Block, P. Towards Automated Building Life Cycle Assessments: A Novel Approach Using Large Language Models and the COMPAS Framework. In *Proceedings of IASS Annual Symposia; International Association for Shell and Spatial Structures (IASS)*: Madrid, Spain, 2024.
40. Gao, Y.; Gan, Y.; Chen, Y.; Chen, Y. Application of large language models to intelligently analyze long construction contract texts. *Constr. Manag. Econ.* **2024**, *43*, 226–242. [[CrossRef](#)]
41. Wong, S.; Zheng, C.; Su, X.; Tang, Y. Construction contract risk identification based on knowledge-augmented language models. *Comput. Ind.* **2024**, *157–158*, 104082. [[CrossRef](#)]
42. He, Y.; Tang, Y.; Chen, T. A Study on Large Language Model-Based Approach for Construction Contract Risk Detection. In *ICBDDM 2024: 2024 International Conference on Big Data and Digital Management*, Shanghai, China, 16–18 August 2024; ACM: New York, NY, USA, 2024; pp. 136–141, ISBN 9798400710278.
43. Kong, F.; Ahn, S. Use of Knowledge Graphs for Construction Safety Management: A Systematic Literature Review. *Information* **2024**, *15*, 390. [[CrossRef](#)]
44. Lee, J.; Ahn, S.; Kim, D.; Kim, D. Performance comparison of retrieval-augmented generation and fine-tuned large language models for construction safety management knowledge retrieval. *Autom. Constr.* **2024**, *168*, 105846. [[CrossRef](#)]
45. Uhm, M.; Kim, J.; Ahn, S.; Jeong, H.; Kim, H. Effectiveness of retrieval augmented generation-based large language models for generating construction safety information. *Autom. Constr.* **2025**, *170*, 105926. [[CrossRef](#)]
46. Karnouskos, S. The Relevance of Large Language Models for Project Management. *IEEE Open J. Ind. Electron. Soc.* **2024**, *5*, 758–768. [[CrossRef](#)]
47. Nyqvist, R.; Peltokorpi, A.; Seppänen, O. Can ChatGPT exceed humans in construction project risk management? *Eng. Constr. Arch. Manag.* **2024**, *31*, 223–243. [[CrossRef](#)]
48. Pokhrel, R.; Parajuli, S. Transforming Structural Engineering: Examining the Opportunities and Risks of ChatGPT and Other Large Language Models. *Int. J. Eng. Technol.* **2023**, *1*, 204–222. [[CrossRef](#)]
49. Zhu, Y.; Zhao, G.; Liu, H.; Sun, D.; He, Y. Refining Bridge Engineering-based Construction Scheme Compliance Review with Advanced Large Language Model Integration. In *Proceedings of the BDIOT 2024: 2024 8th International Conference on Big Data and Internet of Things*, Macau, China, 14–16 September 2024; ACM: New York, NY, USA, 2024; pp. 297–305, ISBN 9798400717529.

50. Chen, X.; Zhang, L. Revolutionizing Bridge Operation and Maintenance with LLM-based Agents: An Overview of Applications and Insights. *arXiv* **2024**, arXiv:2407.10064. [[CrossRef](#)]
51. Ferdousi, R.; Hossain, M.A.; Yang, C.; Saddik, A.E. DefectTwin: When LLM Meets Digital Twin for Railway Defect Inspection. *arXiv* **2024**, arXiv:2409.06725.
52. Gao, Y.; Xiong, G.; Li, H.; Richards, J. Exploring bridge maintenance knowledge graph by leveraging GraphSAGE and text encoding. *Autom. Constr.* **2024**, *166*, 105634. [[CrossRef](#)]
53. Bellan, P.; Dragoni, M.; Ghidini, C. Process Knowledge Extraction and Knowledge Graph Construction Through Prompting: A Quantitative Analysis. In *Proceedings of the SAC '24: 39th ACM/SIGAPP Symposium on Applied Computing, Avila, Spain, 8–12 April 2024*; Hong, J., Park, J.W., Przybyłek, A., Eds.; ACM: New York, NY, USA, 2024; pp. 1634–1641, ISBN 9798400702433.
54. Kommineni, V.K.; König-Ries, B.; Samuel, S. From human experts to machines: An LLM supported approach to ontology and knowledge graph construction. *arXiv* **2024**, arXiv:2403.08345. [[CrossRef](#)]
55. Zhang, B.; Carriero, V.A.; Schreiberhuber, K.; Tsaneva, S.; González, L.S.; Kim, J.; de Berardinis, J. OntoChat: A Framework for Conversational Ontology Engineering using Language Models. *arXiv* **2024**, arXiv:2403.05921. [[CrossRef](#)]
56. Prasad, D.; Pimpude, M.; Alankar, A. Towards Development of Automated Knowledge Maps and Databases for Materials Engineering using Large Language Models. *arXiv* **2024**, arXiv:2402.11323. [[CrossRef](#)]
57. Durmus, D.; Rožanec, J.; Carbonari, A.; Giretti, A.; Turk, Ž.; Grobelnik, M. Potential of Large Language Models for Ontology Development in Construction Domain. In *Proceedings of the Creative Construction Conference 2024, Praha, Czech Republic, 29 June–2 July 2024*; Budapest University of Technology and Economics: Budapest, Hungary, 2024.
58. Wang, B.; Min, S.; Deng, X.; Shen, J.; Wu, Y.; Zettlemoyer, L.; Sun, H. Towards Understanding Chain-of-Thought Prompting: An Empirical Study of What Matters. *arXiv* **2022**, arXiv:2212.10001.
59. Ding, N.; Qin, Y.; Yang, G.; Wei, F.; Yang, Z.; Su, Y.; Hu, S.; Chen, Y.; Chan, C.-M.; Chen, W.; et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nat. Mach. Intell.* **2023**, *5*, 220–235. [[CrossRef](#)]
60. Han, Z.; Gao, C.; Liu, J.; Zhang, J.; Zhang, S.Q. Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey. *arXiv* **2024**, arXiv:2403.14608.
61. Li, Q.; Fu, L.; Zhang, W.; Chen, X.; Yu, J.; Xia, W.; Zhang, W.; Tang, R.; Yu, Y. Adapting Large Language Models for Education: Foundational Capabilities, Potentials, and Challenges. *arXiv* **2024**, arXiv:2401.08664. [[CrossRef](#)]
62. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv* **2022**, arXiv:2201.11903.
63. Gangemi, A.; Catenacci, C.; Ciarmita, M.; Lehmann, J. Modelling Ontology Evaluation and Validation. In *The Semantic Web: Research and Applications*; Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 140–154, ISBN 978-3-540-34544-2.
64. Cristani, M.; Cuel, R. A Survey on Ontology Creation Methodologies. *Int. J. Semant. Web Inf. Syst.* **2005**, *1*, 49–69. [[CrossRef](#)]
65. Wilson, R.S.I.; Goonetillake, J.S.; Indika, W.A.; Ginige, A. A conceptual model for ontology quality assessment. *Semant. Web* **2023**, *14*, 1051–1097. [[CrossRef](#)]
66. Bravo Contreras, M.C.; Hoyos Reyes, L.F.; Reyes Ortiz, J.A. Methodology for ontology design and construction. *Account. Adm.* **2018**, *64*, 134. [[CrossRef](#)]
67. DIN 1076; Ingenieurbauwerke im Zuge von Straßen und Wegen (Engl. DIN 1076—Engineering Structures in Connection with Roads-Inspection and Test): Überwachung und Prüfung. Deutsches Institut für Normung e.V.: Berlin, Germany, 1999.
68. DIN EN 12504-2; Prüfung von Beton in Bauwerken—Teil 2: Zerstörungsfreie Prüfung—Bestimmung der Rückprallzahl; Deutsche Fassung EN 12504-2:2021. Deutsches Institut für Normung e.V.: Berlin, Germany, 2021.
69. DIN EN ISO 9712; German Concrete and Construction Engineering Association. Zerstörungsfreie Prüfung—Qualifizierung und Zertifizierung von Personal der zerstörungsfreien Prüfung (Engl. Non-Destructive Testing—Qualification and Certification of Non-Destructive Testing Personnel). Deutsches Institut für Normung e.V.: Berlin, Germany, 2022.
70. Walther, A.; Hasenstab, A. Zerstörungsfreie Prüfverfahren zur Bestimmung von Materialparametern im Stahl- und Spannbetonbau (Engl. Non-Destructive Testing Methods for the Determination of Material Parameters in Reinforced and Prestressed Concrete Construction). In *Bauphysik-Kalender 2012*; Fouad, N.A., Ed.; Wiley: Hoboken, NJ, USA, 2012; pp. 157–202, ISBN 9783433029862.
71. Horn, K. *Bauwerksanalyse (Engl. Structural Analysis)*; Fraunhofer IRB Verlag; Beuth Verlag GmbH: Stuttgart, Germany; Berlin, Germany; Wien, Austria; Zürich, Switzerland, 2020; ISBN 978-3-8167-9482-0.
72. Taffe, A.; Hillemeier, B.; Walther, A. Verifizierung moderner zerstörungsfreier Prüfverfahren an einem Abbruchbauwerk (engl. Verification of modern non-destructive testing methods on a demolition structure). *Beton Und Stahlbetonbau* **2010**, *105*, 813–820. [[CrossRef](#)]
73. Flohrer, C. Orten der Bewehrung und Messen der Betondeckung. In *Bauphysik-Kalender 2004*; Cziesielski, E., Ed.; Ernst und Sohn: Berlin, Germany, 2004; Chapter C1, Section 2.8; pp. 370–379.
74. Taffe, A. Regelwerke in der Bauwerksdiagnose (Engl. Regulations in Structural Diagnostics). In *Proceedings of the Fachtagung Bauwerksdiagnose 2014*, Berlin, Germany, 13–14 February 2014.

75. Pak, C.-I. Optimierung des Remanenzmagnetismus-Verfahrens zur Stahlbruchortung in Bauwerken—Theoretische Weiterentwicklung (engl. Optimization of the Remanence Magnetism Method for Steel Fracture Detection in Structures—Theoretical Further Development). Ph.D. Thesis, Technische Universität Berlin, Berlin, Germany, 2010.
76. Bundesministerium für Verkehr und Digitale Infrastruktur. Richtlinien für die Erhaltung von Ingenieurbauten (RI-ERH-ING) (Engl. Guidelines for the Maintenance of Engineering Structures RI-ERH-ING). 2020. Available online: <https://www.bast.de/DE/Publikationen/Regelwerke/Ingenieurbau/Erhaltung/RI-ERH-ING.html> (accessed on 27 January 2025).
77. Merkblatt B 02: Merkblatt zur Zerstörungsfreien Betondeckungsmessung und Bewehrungsartung an Stahl- und Spannbetonbauteilen (Engl. Code of Practice for Non-Destructive Concrete Cover Measurement and Reinforcement Location on Reinforced and Prestressed Concrete Components (Code of Practice B02)); Deutsche Gesellschaft für Zerstörungsfreie Prüfung: Berlin, Germany, 2014.
78. Merkblatt B 03: Elektrochemische Potentialmessungen zur Detektion von Bewehrungsstahlkorrosion (Engl. Electrochemical Potential Measurements for the Detection of Reinforcement Steel Corrosion—Code of Practice B03); Deutsche Gesellschaft für zerstörungsfreie Prüfung: Berlin, Germany, 2021.
79. Merkblatt B 04: Ultraschallverfahren zur Zerstörungsfreien Prüfung im Bauwesen (Engl. Ultrasonic Methods for Non-Destructive Testing in the Construction Industry—Code of Practice B04); Deutsche Gesellschaft für zerstörungsfreie Prüfung: Berlin, Germany, 2018.
80. Merkblatt B 10: Merkblatt über das Radarverfahren zur Zerstörungsfreien Prüfung im Bauwesen (Engl. Leaflet on the Radar Method for Non-Destructive Testing in the Construction Industry—Code of Practice B10); Deutsche Gesellschaft für zerstörungsfreie Prüfung: Berlin, Germany, 2008.
81. Merkblatt B 12: Korrosionsmonitoring bei Stahl- und Spannbetonbauwerken (engl. Corrosion Monitoring for Reinforced and Prestressed Concrete Structures—Code of Practice B12); Deutsche Gesellschaft für zerstörungsfreie Prüfung: Berlin, Germany, 2018.
82. Roumeliotis, K.I.; Tselikas, N.D. ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet* **2023**, *15*, 192. [CrossRef]
83. Alto, V. *Modern Generative AI with ChatGPT and OpenAI Models: Leverage the Capabilities of OpenAI's LLM for Productivity and Innovation with GPT3 and GPT4*; Packt Publishing: Birmingham, UK, 2023; ISBN 9781805122838.
84. Sider: ChatGPT Sidebar + GPT-4o, Claude 3.5, Gemini 1.5 & AI Tools. Available online: <https://sider.ai/en> (accessed on 8 January 2025).
85. Auger, T.; Saroyan, E. Overview of the OpenAI APIs. In *Generative AI for Web Development*; Auger, T., Saroyan, E., Eds.; Apress: Berkeley, CA, USA, 2024; pp. 87–116, ISBN 979-8-8688-0884-5.
86. Strangmann, T.; Purucker, L.; Franke, J.K.H.; Rapant, I.; Ferreira, F.; Hutter, F. Transfer Learning for Finetuning Large Language Models. *arXiv* **2024**, arXiv:2411.01195. [CrossRef]
87. Lu, W.; Luu, R.K.; Buehler, M.J. Fine-tuning large language models for domain adaptation: Exploration of training strategies, scaling, model merging and synergistic capabilities. *arXiv* **2024**, arXiv:2409.03444. [CrossRef]
88. Chen, B.; Zhang, Z.; Langrené, N.; Zhu, S. Unleashing the potential of prompt engineering in Large Language Models: A comprehensive review. *arXiv* **2023**, arXiv:2310.14735. [CrossRef]
89. Amatriain, X. Prompt Design and Engineering: Introduction and Advanced Methods. *arXiv* **2024**, arXiv:2401.14423. [CrossRef]
90. van Buren, D. Guided scenarios with simulated expert personae: A remarkable strategy to perform cognitive work. *arXiv* **2023**, arXiv:2306.03104. [CrossRef]
91. Hamdan, A.-H.; Kozak, T. Building Material Definitions (BMAT). Available online: <https://alhakam.github.io/bmat/> (accessed on 29 January 2024).
92. Noy, N.F.; McGuinness, D.I. Ontology Development 101: A Guide to Creating Your First Ontology. *Comput. Sci.* **2001**.

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.