SMART ONTOLOGY CONSTRUCTION: COMBINING HUMAN EXPERTISE WITH LLMS FOR SMARTER BUSINESS TAXONOMIES

by

Ramanathan Rajamani

DISSERTATION

Presented to the Swiss School of Business and Management, Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

MAY, 2025

SMART ONTOLOGY CONSTRUCTION: COMBINING HUMAN EXPERTISE WITH LLMS FOR SMARTER BUSINESS TAXONOMIES

by

Ramanathan Rajamani

Supervised by

Dr Anuja Shukla Phd

APPROVED BY

____

Dissertation chair – Dr. Gualdino Cardoso

RECEIVED/APPROVED BY:

____

Admissions Director

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Anuja Shukla, for her invaluable guidance, support, and insightful feedback throughout this research. Her expertise in Data Analytics and erudition in ensuring a tight research temper was instrumental in shaping the direction and quality of this dissertation.

ABSTRACT

SMART ONTOLOGY CONSTRUCTION: COMBINING HUMAN EXPERTISE WITH LLMS FOR SMARTER BUSINESS TAXONOMIES

Ramanathan Rajamani
2025

Dissertation Chair: Dr. Gualdino Cardoso
Co-Chair: Dr. Sasa Petar

Ontologies are a fundamental piece in the development of semantic web technologies, artificial intelligence applications and knowledge representation systems. While the Large Language Models (LLMs) are very capable of automatically generating ontologies by detecting concepts, taxonomy structures and relationships, they still rely on human input because of the limitations of logical reasoning, hierarchical structuring, and contextual accuracy. To improve the quality, scalability, and domain relevance of LLM assisted ontology construction this research presents a Human in the Loop (HITL) framework. The study employs OSHA accident and injury data, data enhancement, exploratory data analysis (EDA), and structured prompt engineering with GPT-4 to automate ontology generation, expert review to fix logical inconsistencies, resolve ambiguities, and remove redundancy. Ontology quality, accuracy, completeness, relevance, and consistency are measured and the automated model starts with 0.69 accuracy, 2.46 relationships per record completeness, 0.78 relevance, which are all improved upon in human refinement. The improvements are validated quantitatively through paired t-tests showing statistically significant gains in accuracy, consistency, and relevance, and qualitatively through a

structured expert questionnaire involving 15 domain experts across three evaluation rounds. This study also shows the practical benefit of domain specific taxonomies that have been shown to reduce operational decision latency, enhance machine learning prediction precision, and accelerate and affordably transform the digital. Taxonomy development is a core strategic capability that organizations with a mature taxonomy benefit from accelerated cloud migration, reduced integration costs, and significant annual savings. The research integrates human expertise and LLM-driven automation to provide a robust, data driven framework for enhancing decision making, improving operational efficiency, and creating a competitive advantage in the digital frontier as it lays the groundwork for further advances in reasoning capabilities, bias mitigation, and prompt engineering in the design of ontology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviations	Full Form
LLMs	Large Language Models
HITL	Human-In-The-Loop
EDA	Exploratory Data Analysis
NLP	Natural Language Processing
ML	Machine Learning
RAG	Retrieval Augmented Generation
HRM	Human Resource Management
GNNs	Graph Neural Networks
XAI	Explainable AI
KG	Knowledge Graph
SMS	Systematic Mapping Study
OEM	Ontology Engineering Methodology
SAR	Search and Rescue
HBCP	Human Behaviour-Change Project
OBDA	Ontology-Based Data Access

CHAPTER I:

INTRODUCTION

1.1 Introduction

This section gives a description of what is meant by the construction of ontologies, the importance of ontology construction and its development. It discusses how ontologies, as fixed structures, are used for organizing information in different topic areas (Ashburner et al., 2000). The section also explores the shift from the conventional, time-consuming technique of ontology development to the use of Artificial Intelligence (AI) informed methodologies, together with the advantages and drawbacks of each (Asim et al., 2018; Boccacci et al., 2015). Furthermore, it explores different uses of ontologies in today's world, including health care, finance, and artificial intelligence. Given that this section is laying the groundwork, its goal is to demonstrate why constructing ontology is crucial for the effective organization of data and to support higher-level computations.

Ontology and Its Importance

In Information Science and AI, ontology models a domain or topic of knowledge. It defines concepts, their properties, and their relationships. A successful ontology enables robots and systems gather and process information to attain 'semantic identity' with the human approach (Guarino et al., 2009). This organized form is helpful for processing unstructured data, integrating systems, and improving decision-making. To connect human and machine thought, ontologies are needed (Guizzardi & Guarino, 2024). Ontologies underpin semantic web technologies, knowledge graphs, and smart systems in health care, finance, education, and e-business (Radain, 2022). They improve data integration and retrieval by providing a shared vocabulary and conventional relationships. They are necessary because vast and complex data collections must be retrievable, intelligible, and useable.

Structuring and formalizing the domain knowledge is an important role of ontologies; it is the backbone of semantic technologies, artificial intelligence and enterprise-level data organization. In the context of business strategy and digital transformation, structured ontological hierarchies—domain-specific taxonomies have been proven to be useful in improving decision speed, system interoperability, and data-driven automation. For example, case studies of Fortune 500 have shown domain-specific taxonomies to reduce operational decision latency by 37 percent (Harvard Business Review, 2022). Adaptive taxonomies further enhance AI capabilities to the extent of up to 89 % more accurate machine learning predictions in supply chain optimization scenarios (MIT Sloan Management Review, 2023). Organizations that have matured their taxonomy systems realize 2.1x faster cloud migration and 45% lower integration costs, and estimated annual savings of $4.3 million for mid-sized enterprises by just improving search functionalities and recommendation systems (Gartner, 2024). Furthermore, ontology-guided AI systems have increased decision accuracy by 42% on manufacturing use cases, and frameworks such as SMART (SMART: Human in the loop refinement of LLM generated ontologies) have up to 89% consensus between automated schema outputs and expert-curated structure (X. Chen & Watanabe, 2023). By providing these insights, ontology development is positioned as a business strategic imperative that directly enables business agility, predictive accuracy, and scalable knowledge management.

Knowledge integration and sharing via ontologies has gained prominence in the last decade. Due to the widespread belief that computers cannot comprehend human language, ontology quickly became a byword for the semantic web's promised answer to common issues (Biemann, 2005). Recognition of ontology for knowledge representation and information retrieval and extraction has spread across many academic disciplines (Z. Li & Ramani, 2007). Conceptually, "ontology" refers to entities and the connections between them. All of these databases have its own set of implicit or explicit ideas. There is a systematic approach to study things that already exist, and ontology gives a way to categorize such objects (Jain & Mishra, 2016). A knowledge-based system makes use of ontology, which is a kind of knowledge. Any knowledge base system may access the ontology, which is a data definition stored into the database. Web document retrieval heavily employs ontology, as does information retrieval via keyword-based searching, which might provide some false-matched results due to the fact that extraction is not yet an option (Jain & Mishra, 2016). This information is only available when the keywords match, not when the meaning is known. Based on a common conceptual domain model, ontology facilitates knowledge management by linking various technological systems for information storage, retrieval, and exchange (Osman et al., 2022). Representing knowledge and ideas is where ontology comes in. It lays forth methods for handling data that is not organized. Constructing ontologies is a computationally feasible method of idea organization that facilitates concept transferability between actors and levels of abstraction.

Ontologies for knowledge integration and exchange have gained in relevance over the past decade. Ontology soon became a byword for the semantic web's promised solution to common problems due to the idea that computers cannot understand human language (Biemann, 2005). Many academic fields recognize ontology for knowledge representation, retrieval, and extraction (Z. Li & Ramani, 2007). Conceptually, "ontology" means entities and their relationships. Each database has implicit or explicit ideas. Ontology categorizes Figure existing objects for systematic examination (Jain & Mishra, 2016). Ontology is used in knowledge-based systems. The ontology, a database data specification, is accessible to any knowledge base system. Since extraction is not yet possible, keyword-based searching and web document retrieval largely use ontology, which may yield false-matched results (Jain & Mishra, 2016). This information is only available when keywords match, not when meaning is known. Ontology connects technical systems for information storage, retrieval, and exchange using a shared conceptual domain model (Osman et al., 2022). Ontology represents knowledge and ideas. Methods for processing unorganized data are provided. Ontologies enable concept transfer between actors and abstraction levels and are computationally possible.

Evolution of Ontology Construction: From Manual to AI-Assisted Methods

Over time, ontology creation has moved from manual to AI-assisted methods. Initial ontologies required domain specialists to manually define concepts, relationships, and characteristics. This method was slow and unproductive since it relied on staff opinions and estimates (Faria et al., 2014). However, manual ontology development created the first knowledge representations useful for organizing medical, biological, and engineering information. AI—especially NLP and ML—has simplified ontology construction. AI algorithms recognize concepts, deduce relationships, and categorize text or database data (Baviskar et al., 2021). “Large Language Models” (LLM) like GPT can construct ontologies without human input, advancing this idea (Olga Perera, 2024). These approaches also improve construction efficiency and ontology precision and flexibility across domains. Due to the contextual nature of the problem, this study demonstrated that AI-generated ontologies still need human correction and guidance.

Intelligent systems now automate ontology building. Creating ontologies involved coding field-specific ideas, relations, and attributes. This lengthy and uneven process relied on professionals' eyes and judgment to examine data. Manual ontology construction before AI development propulsion was difficult for designing early medical, biological, and engineering knowledge representation systems (Yun et al., 2011). Ontology development is faster, simpler, and better thanks to ML and NLP AI (Solanki, 2024). AI systems correlate, organize, and detect concepts in complex text or database data. Ontologies created by LLMs like GPT with little human interaction have progressed this field (Aggarwal et al., 2024). This speeds up building and improves ontology flexibility and precision. AI-generated ontologies' context-reactivity causes various issues that require human involvement. Thus, much progress has been accomplished (Rane et al., 2024).

The "digital revolution" in cloud computing will replace existing IT paradigms in the next decade. The cloud market is young, but many reasons could lead to monopolistic or anti-competitive practices (S. Song, 2017). Some suppliers negotiate strangely or exclusively and refuse to disclose technical details about compatible goods. Pricing and monopolistic behaviour reduce competition and innovation. Competition legislation and other laws hinder cloud computing service providers' ability to compete. Concentration laws may affect CC sector market concentration efforts. Interoperability is a major competition law and mergers issue. This principle is crucial in cloud computing since it affects transparency, competitiveness, standardization, and IP rights. According to Song (2017) and Bornico and Walden (2011)Competition law continues to prevent dominant market firms from abusing their position, even though it isn't keeping up with technological advances. Competition law reviews and investigations of software and hardware platform monopolies are common. Cloud computing points of sale may be regulated.

However, Ontology creation remains problematic in AI-supported approaches. Domain knowledge and context-specific interpretation and LLMs may not cause semantic variance with diverse data kinds. Thus, knowledge participation is still possible and required. AI outputs are more accurate, context-relevant, and able to meet domain-specific needs, so humans examine, enhance, and approve them. The future of building ontologies will include human intelligence and the separation of automation into the parts of the problem that computers can best solve. Both will create solid, scalable, and contextually accurate ontologies for the intended application.

Applications of Ontologies in Modern Systems

Ontologies are critical in the current generation systems because they offer a structural way of implementing, assimilating as well as analyzing data. They help machines and people to have a similar interpretation of information content that makes them vital tools in diverse fields (Tiwari et al., 2011). One of their main uses is in the Semantic Web to assist in the organization of web data so that improved and more relevant results can be returned. Technologies such as RDF and OWL use only ontologies to construct a set of related datasets which form a knowledge graph used by search engines such as Google (Verma et al., 2023). In healthcare, we can find several examples of ontologies like SNOMED CT and Gene Ontology which are used for medical diagnosis, medical researches and for discovery of some new drugs (El-Sappagh et al., 2018). They work to normalize medical language, promote interoperability between organizations and improve the management and analysis of data by correlating patient information with information that physicians need at point of care (Grannis et al., 2019). Similarly, in e-business, ontologies enhance recommendation system since product attributes/ characteristics and customer preferences for the products are well arranged, creating suitable recommendations for the clients.

Another is in AI where ontologies contribute to; machine reasoning, natural language processing, and autonomous systems (Beg, 2024). They allow AI models to identify patterns, reason and deduce new information and domain expertise. This is because ontologies give AI systems a way of organizing and allocating the available knowledge to make them smarter, more contextually sensitive and better fitted to solve real-world problems. In an era of rapid technological progression and information exchange, the role of structuring, storing, and processing data becomes increasingly important (Peter C. Verhoef, Thijs Broekhuizen, Yakov Bart, Abhi Bhattacharya, John Qi Dong, Nicolai

Fabian, 2021; Storey, 2019). Ontologies, as a tool for representing knowledge in a systematized form, hold a special place in modern information systems (Jurisica et al., 1999). They allow for the creation of standardized data structures that can unify disparate pieces of information into a coherent representation. This tool has long become an integral part of many sectors, from AI to medicine. Ontologies help professionals work effectively with complex data, making it accessible and understandable. However, like any powerful tool, ontologies have their limitations and specific issues. Interactions between different research groups, the creation of universal ontologies, the integration of the human factor, and new risks in the realm of information security – all these make the issues of applying and developing ontologies pertinent. In this research, they will consider the key aspects of using ontologies in modern information systems, their capabilities and limitations, and also propose alternative approaches to data structuring and processing (Pospelov, 2023).

Medical workers diagnose and cure disorders in healthcare. Healthcare providers, researchers, and technologists collaborate to provide high-quality, affordable care (D. Cortes et al., 2019). They generate a lot of data from many sources to enhance communication between doctors and patients, treatment decisions, and diagnostic accuracy. They must organize critical data so they can find it when needed (Clemente et al., 2022; Haque et al., 2022).

Besides these, ontologies are becoming the foundation for future technology trends like IoT, smart cities, and autonomous systems. Ontologies assist smart homes and industrial automation make smart judgements by structuring data transmission between devices (Ryabinin et al., 2019). Ontologies assist build smart cities by collecting data from multiple sectors like transport, energy, and public utilities (Komninos et al., 2016). In autonomous systems like self-driving cars, ontologies provide a semantic perspective of the world, improving situation awareness and time-bound judgements (Ignatious et al., 2023). Ontologies are utilized in science and education. They let researchers to tag and publish datasets and data files for standardization and repeatability. The Environment Ontology (ENVO) helps environmental scientists categorize and connect systems data for research and policy. Ontologies are used to construct intelligent learning environments that tailor information to a learner's goals, attributes, and achievements.

As technology progresses and challenges grow more interdisciplinary, ontologies become almost essential to solution seeking procedures, ensuring their relevance and applicability. Ontologies are considered to be crucial to the design of unique, problem-solving intelligent systems by providing a common platform for data understanding.

The Role of Large Language Models (LLMs) in ontology generation

The author examines LLMs in modern AI technologies and ontology building in this section. The article introduces LLMs and their benefits for natural language understanding and generation. The section shows how such models can simulate information extraction, concept recognition, and knowledge organization and presentation. The goal of implementing LLMs into ontology construction is to make the process more efficient and scalable in comparison to human workers.

It also examines LLMs' biases, mistakes, and need for context to interpret, but it emphasizes the significance of human intelligence to improve them. This section explains how LLMs' properties contribute to NOTION and AI knowledge management, as well as their uses and problems.

Current examples show that LLMs are not highly intelligent, even though AI may govern the world. Although amazing and adaptive, these technologies can interpret human language and answer complex questions (Mannuru et al., 2023; Yan et al., 2023). Concepts can be defined and natural language sentences converted to OWL. For the purpose of creating ontologies, LLMs compile information representations and definitions of terms in both formal and natural languages (Booshehri et al., 2021; Neuhaus, 2023). In contrast to ontologies, which are tools that subject-matter specialists directly create and maintain, LLMs are general-purpose tools (Boccacci et al., 2015)

Overview of LLM Capabilities in Natural Language Understanding

The recent advance of PLMs like BERT Devlin *et al.* (2019) and GPT Radford (2018) families has fairly dominated "Natural Language Processing" (NLP) benchmarks.

An entirely new paradigm in NLP has emerged as a result of these massive PLMs. As an example, consider the classification problem p(y/x), which involves assigning a label y to textual input x: Simple, manually created features for x are typically used in classic statistical NLP techniques for y, after which a classifier (like SVM) is trained. Cortes and Vapnik (1995), (logistic regression) Models for Deep Learning (DL) absorb the hidden feature representation using an ANN (LeCun et al., 2015) with its classifying function. Keep in mind that every "natural language processing" (NLP) work requires learning the latent representation from scratch, and that sometimes quality of the latent feature representation is limited by the amount of training data. It is reasonable to assume that all NLP activities share linguistic subtleties, and that we might thus acquire a generic representation of latent features from one generic job and distribute it across other NLP tasks (Gatt & Krahmer, 2018). Being a generic problem that can be pre-trained using a huge amount of naturally occurring text, "pre-trained language models" allow models to learn to predict the next word given prior words. This is how language modelling works (Hadi et al., 2023). Actually, the advent of PLMs marked the beginning of the most recent paradigm change. Researchers are currently leveraging pre-existing PLMs for numerous NLP activities. They do this in one of two ways: by recasting the problem as one of text creation and employing PLMs to find the right solution, or by honing in on the specific job at hand and giving the PLMs specific instructions to complete it. New state-of-the-art performances have continuously been generated by these three PLM-based paradigm advances (Barak & Mokfi, 2019; Biloshchytskyi et al., 2017).

Text classification and summarization, sentiment analysis, and high-performing quizzes are LLM exceptions for NLU. LLMs can summarize large reports or assess client feedback to determine its tone. Thus, they are effective at interpreting contextual meaning in language translation and conversational AI intent identification (Hadi et al., 2023). Due to their adaptability, LLMs are essential for content production and client service.

Language is essential to relationships and the ability of humans and robots to communicate. Progressive models are needed because machines must do language-oriented tasks like translation, summarization, information search, dialogues, etc. Transformers, computational improvements, and massive training data sets have advanced language models in recent years (Chernyavskiy et al., 2021). Evolutionary advancements have allowed LLMs to do human-like tasks (Wu and Lode, 2020; Brown *et al.*, 2020). Big language models, the latest NLP technology, can reason and generate text in coherent conversational style (Agüera Y Arcas, 2022) and work in many fields.

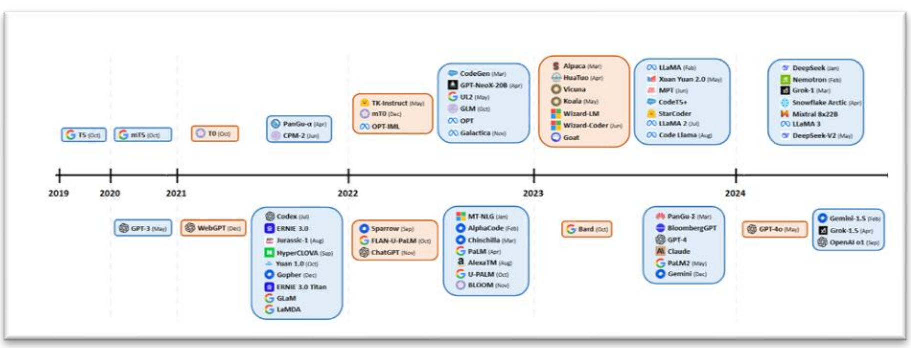

Figure 1.1: Chronological display of LLM releases.

Source: **-** (Naveed et al., 2023)

Figure 1.1: Timeline for the publication of LLM One set of playing cards is for "instruction-tuned" models, and the other set is for "pre-trained" ones. There are half-open-

source models and half-closed-source models here. A change in the field and trends within NLP research is indicated by the tendency towards more instruction-tuned and open-source approaches.

Historical NLP development started with statistical modelling, progressed to neural modelling, PLMs, and finally LLMs. In contrast to standard LM, which uses supervised scenarios to train task-specific models, PLMs are pre-trained unsupervised using a large text corpus (Devlin *et al.*, 2019; Peters *et al.*, 2018; Lewis *et al.*, 2020) In accordance with the developers' intention, they learned a general representation usable in multiple NLP tasks. Once tuned for downstream tasks, PLMs outperform traditional LM in terms of as a downstream task. The migration of PLMs to LLMs has been caused by the massive expansion of model parameters (tens to hundreds of billion), since the larger PLMs yield more performance advantages (Raffel et al., 2020) training corpus (a large number of GBs and TBs) (Raffel *et al.*, 2020; Xue *et al.*, 2021). Since then, a large number of LLMs have been put up in the literature (Raffel *et al.*, (2020); Xue *et al.*, (2021); Zhang *et al.*, (2021); Brown *et al.*, (2020); Le Scao *et al.*, (2022)).

Additionally, LLMs can be scaled up to solve other tough natural language understanding challenges. Their zero-shot, few-shot, and transfer learning methods allow them to solve problems without retraining or labelling enormous datasets. It is useful in changing fields like medicine, law, and communication. As with other LLMs, they are vulnerable to bias in training data and struggle with domain specificity of definitions. LLMs are being used more as research improves them, reduces biases, and improves interpretability. These models enable more natural and intelligent human computer interfaces, benefiting many sectors.

LLMs in Automating Ontology Development

LLMs have recently proven to be useful for automating the process of ontology creation by reducing the tasks which used to be done manually to a mere formality (Lo et al., 2024). Ontology creation involves the identification of entities, definition of properties, and definition of relations or relationships between them and this is always a tiresome activity (González-Eras et al., 2022). Compared to simpler ML models, LLMs are more sophisticated and can process massive amounts of unstructured textual input, finding related concepts and their relationships and then transforming them into organized forms (Yuan et al., 2024). Consequently, LLMs effectively facilitate the development of ontologies due to the ability of the LLM to automatically detect and categorize domain-specific knowledge.

Working with academic, report, and Web data to develop domain ontologies is another key advantage of LLMs in this domain (Ling et al., 2023). They can extract medical terms from research articles to create health care ontologies or product features from e-commerce data to create recommendation ontologies (Alsobhi & Amare, 2022). LLMs are also open to progressive changes based on domain specialist feedback to ensure that generated ontologies satisfy specific needs and are context-sensitive. These scholars have discreetly adopted LLM automation and flexibility to further ontology development in numerous domains.

KGs make it efficient to collect, infer, retrieve, and analyze structured data. In a graph structure, KGs mimic the actual world by using nodes to represent entities and edges to indicate their connections (X. L. Dong, 2023). The use of KGs in NLP has recently shown promise for improving LLM efficiency (Agrawal *et al.*, 2023; Galkin *et al.*, 2024; Pan *et al.*, 2024). LLMs are known to have ‘hallucinations’ when the system feeds them false information. (Xu et al., 2024). Retrieval Augmented Generation (RAG), where graph structure is encoded as textual information as input to the LLM, has been demonstrated to help alleviate this problem (Edge *et al.*, (2024); Jiang *et al.*, (2023); Sen, Mavadia and Saffari, (2023); Wu *et al.*, (2023)). We also know that KGs improve LLM reasoning. (Luo *et al.*, 2024; Wen, Wang and Sun, 2023).

However, developing LLMs for ontology building automation is difficult, as shown below. Lack of precision for domain-specific tasks, transmitting training data bias to outcomes, and inability to explain generative model outputs are problems. Many domain-specific criteria for ontology validation and improvement must be evaluated by humans, even while LLMs can recognize and categorize ideas. In other words, LLMs are doing more regular data processing while domain specialists provide context-aware accuracy. This approach also accelerates ontology development and ensures the production of a sound, contextually relevant knowledge structure for various use.

Key Features and Limitations of LLMs in Ontology Construction

Ontology building benefits from LLMs due to their key properties. One of Hercules' main strengths is its capacity to automatically discover concepts, entities, and relationships in enormous amounts of unstructured text (Ibrahim et al., 2024; Schilling-Wilhelmi et al., 2024). Compared to the manual process, this saves time and effort. LLMs can generate area-specific ontologies from several datasets (Capellini et al., 2024). They are also semantically skilled enough to find a relationship between two things in the source. They help automate ontology generation for technological and information-intensive domains including health, business and finance, and ecommerce.

Despite their benefits, LLMs have insurmountable drawbacks. I found this to be true with GA in (Hassanin & Moustafa, 2024). Unbalanced or inaccurate training data can affect ontologies. The ontology may be misrepresented. Due to their rule-based nature, they may misread or oversimplify relationships without domain-specific expertise. Users cannot easily contest or interpret the rationale behind LLM outcomes since processes are unclear.

Computer software contains a BIM model of assets' functional and physical features (Solihin & Eastman, 2015). BIM improves project delivery, resource use, and value and objectives (Borrmann et al., 2018). However, as BIM technology is adopted, compliance with complex and ever-changing standards and procedures throughout design and construction becomes more difficult (Nawari, 2012).

It was found that manual compliance checking approaches are not only time consuming, but also inaccurate (Eastman *et al.*, 2009; Tan, Hammad and Fazio, 2010). This inefficiency and susceptibility to errors are against the lean management and engineering concept. For that reason, traditional techniques may not be sufficient to achieve project objectives and deal with a vast amount of data in large-scale projects that has increasingly stringent regulation requirements to address (Ismail et al., 2017).

Many scholars worldwide have studied and tried different techniques to promoting AEC compliance checking due to globalization. They developed ARC systems for Singapore's CORENET, Norway's HITOS, Australia's Building Codes Board, the International Code Council, and the US. (Eastman *et al.*, 2009; Building, Modeling and Check, 2024). At its core, the rule-checking process in the GSA project consisted of four stages: rule extraction (the process of converting rules from natural language to a computer-readable format), building model preparation (the gathering of all the necessary information for the checking procedure), rule execution (the process of using the computer-processable rules to verify the prepared model), and reporting the checking results 2009 (Eastman et al., 2009).

However, LLMs' shortcomings in ontology construction require more steps and a dual strategy that integrates automated and human participation. Automated knowledge management and LLMs can swiftly acquire and organize massive volumes of data, but only topic specialists can keep the ontologies industry-specific and correct. The ontology improves and bias is reduced by clarifying relationships that may be unclear when implemented alone. When LLMs and human understanding are combined, the ontology generation process becomes stronger, more flexible, and the knowledge structures more accurate and inclusive. As the aforesaid technologies evolve, better ways to integrate them together will lead to faster and better ontology building for different domains and fields.

Human-in-the-Loop (HITL) intervention

In this section, the concept of Human-in-the-Loop (HITL) is described as a cooperation between humans and artificial intelligence and machine learning. HITL is first defined and its importance in improving the usability, credibility, and legal accountability of the AI-driven results is stressed. The section explains how HITL processes use human decision-making and situational awareness to mitigate the disadvantages that automatons have such as developing or inheriting prejudice and uncertainty.

In addition, it examines how HITL is put to use by different companies for things like improving data models, enhancing decision-making, and ensuring quality in high-risk sectors. In this section, the need to emphasize HITL when considering the combination of human supervision and AI capacities is described.

Concept and Principles of HITL

HITL is an AI and ML approach that involves human skills into the loop of decision making (Mosqueira-Rey et al., 2023). If in fully autonomous systems AI solely controls all aspects and functions, then in HITL human feedback, supervision and decisions enhance the output accuracy, reliability and context (Kinney et al., 2024). The main concept of HITL is to utilize the advantage of AI; the former is good at dealing with massive amounts of data and following through routine tasks (Siemens et al., 2022), while the latter are capable of providing expertise, insights, and problem-solving skills to counter AI’s weakness, namely its inability to handle some aspects of a job with much creativity and sensitivity.

HITL loops usually include human input and AI system response (Mollick et al., 2024). Supervised machine learning requires humans to label data for the model to learn accurate patterns. During deployment, human input can fine-tune AI prediction or output in real time. This iteration, feedback, and correction process increases model quality and protects human trust in artificial systems since humans make decisions (Alijoyo et al., 2024). HITL excels in ethics, legal, and finance training that requires precision or decision-making. HITL mitigates biased, erroneous, or unethical ideas that the automated system may miss by incorporating human intelligence into decision-making (Gómez-Carmona et al., 2024). HITL is a basis for developing dependable and safe AI applications since it ensures that artificial intelligence is effective not only technically but also in accordance with our and others' values and the demand for solutions in certain sectors.

ML can create computational models that power user-facing apps (Shrestha et al., 2021). While end-users are often the application's domain experts, they rarely participate in its creation. Due to the challenges of applying machine-learning techniques to common issues, they are generally used by experts in 2023 (Morandini et al., 2023). Traditional applied ML practitioners acquire data, select features to represent it, convert and pre-process it, choose a learning algorithm and representation, tweak the algorithm's parameters, and evaluate the model's quality. This review often promotes further method refinement. Practitioners usually limit end-user participation to data provision, domain-related question responding, and learning model feedback. Because of this, the design process is extensive and asynchronous, and end users have little control over the final models.

Biochemists and machine-learning experts clustered low-level protein structures to create a protein taxonomy (Caruana et al., 2006). The project manager shared their experiences at the 2013 Workshop on Interactive Machine Learning for Intelligent User Interfaces (Amershi et al., 2013). To visualize the results, practitioners clustered protein structures and then reviewed them with the biochemist. After reviewing the results, biochemists may say, “These two proteins belong/do not belong to the same cluster” or “There should be more members in this cluster”. After each meeting, practitioners adjusted clustering parameters to stay within restrictions and computed clusters for the next meeting. Tired of this slow process, Caruana and colleagues devised learning algorithms that let individuals interactively traverse the clustering space and add new restrictions (Caruana et al., 2006). This lets people simply switch clustering in one sitting.

As a result, DL has achieved impressive results in a number of domains, including medical applications, ITS enhancements, picture and video decoding, and the comprehension and processing of natural language and speech (S. Dong et al., 2021). This achievement is a result of using more extensive models with several parameters, which provide greater flexibility and descriptive power (Brutzkus & Globerson, 2019). Nevertheless, it is to be also noted that the success of DL depends on a large number of labeled training samples (B. Zhou et al., 2014). The data grows in size at a far slower rate than the number of model parameters, making it a laborious and complicated procedure to acquire and categorize such data. To tackle this difficulty, new datasets are being created, the velocity of model deployment is being increased, and the cost of data annotation is being reduced (J. Li *et al.*, (2021); Zhao *et al.*, (2020); Shen *et al.*, (2021); S. Li *et al.*, (2021)). In addition, pre-trained models for DL and transfer learning techniques such Transformers Vaswani *et* al. (2017), BERT Devlin *et al.* (2019), and GPT (Radford, 2018) are found to give good results. Although the produced data is helpful for getting the model started, further individualized data labelling and updates are needed to create a high-precision, useable model. To address the data scarcity issue, academics have proposed weak supervision approaches like few-shot learning as ways to learn from small data sets (Habermann *et al.*, 2020; Wang *et al.*, 2020; Jia *et al.*, 2021).

Thus, DL has improved image and video decoding, natural language and speech processing, medical applications, and ITS (S. Dong et al., 2021). This achievement is due to larger models with many parameters that allow more description freedom (Brutzkus & Globerson, 2019). DL requires several labelled training samples to succeed (B. Zhou et al., 2014). Data grows much slower than model parameters, making acquisition and categorization laborious and complicated. To address this issue, new datasets are being developed, model deployment is being accelerated, and data annotation costs are being decreased (J. Li *et al.*, (2021); Zhao *et al.*, (2020); Shen *et al.*, (2021); S. Li *et al.*, (2021)). Transfer learning approaches like Transformers and DL models like pre-trained models Vaswani *et* al. (2017), BERT Devlin *et al.* (2019), and GPT (Radford, 2018) also perform well. Though the produced data is useful for getting the model off the ground, accurate results need custom data labelling and improvements. Researchers have created poor supervision and few-shot learning strategies to handle data shortage (Habermann *et al.*, 2020; Wang *et al.*, 2020; Jia *et al.*, 2021).

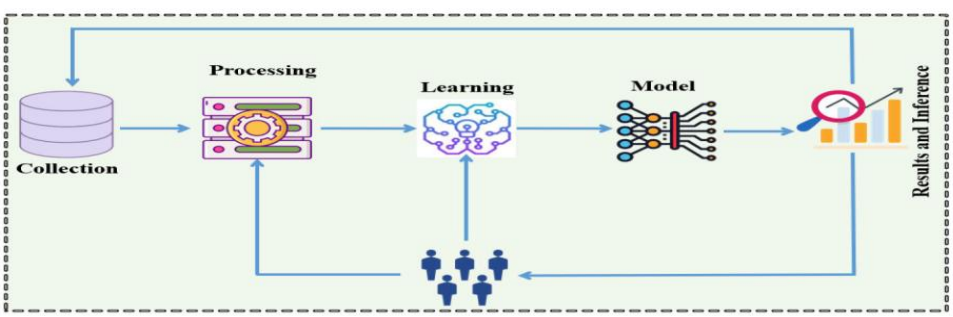

Figure 1.2: Human-in-the-loop learning framework.

Source: - (Kumar et al., 2024)

To tackle sparse data issues, it has been helpful to include past information into the learning architecture. Machines may learn from existing knowledge bases with the assistance of human intelligence and expertise (Diligenti et al., 2017). In fields like clinical illness diagnosis, where data accessibility is sometimes limited, researchers have been incorporating existing information into the acquisition of new knowledge with the belief that knowledge is power (Chen, Leng and Labi, 2020; Lin, Pintea and van Gemert, 2020; Hartmann, Shiller and Azaria, 2019). Employing knowledge from pre-training can improve performance and cope with data scarcity issues adequately, (Zhang *et al.*, 2018; Holzinger *et al.*, 2019; Zhuang *et al.*, 2017). A common ML framework that uses HITL learning is shown in Figure 1.2. It has three parts: cognitive functions, which include data pre-processing, data modelling, and process control, all with goal of improving performance (Xin et al., 2018). However, as ML model outcomes and performance are not always predictable, it is not always easy to determine which part of human-machine interaction yields the most beneficial learning outcomes. Altogether, HITL adheres to three critical principles that revolve around Automation when dealing with complex systems of AI and the fact that the systems require human intervention. Linked to decision making, HITL does away with the demerits of automatic decision-making systems like; prejudices and absence of contextuality by incorporating human elements in the procedures. This gives the general public confidence in the AI systems and also empowers humans to accept outputs that the AI systems give while checking whether they meet the set goals and ethics. As mobility of AI technologies advances, HITL remains to be an important model for designing smart systems that are able to learn from human input, precise and ethical.

Role of Human Expertise in AI Systems

Human expertise plays a pivotal role in AI systems, serving as the critical link between raw machine outputs and meaningful, context-aware decision-making (Lepri et al., 2021). As impressive as AI is at sifting through mountains of data in search of patterns, it frequently falls short when compared to humans when it comes to providing nuanced understanding and judgement. Human experts contribute by defining problem scopes, curating high-quality training datasets, and validating the results generated by AI systems (Lebovitz et al., 2021). Such engagements guarantee that the AI models meet the standards of the various domains and that the existing ethical concerns are met courtesy of the following. Furthermore, humans are involved to overcome the weaknesses, like overstated or understated results, and sample prejudices that come with it. Most of these can be fixed by experts in order to ensure that the AI system is supplying fair and accurate results. In dynamic scenarios where context, ethics, or cultural understanding are essential, human input guides AI systems to make decisions that are both precise and socially responsible (Heyder et al., 2023). By combining computational efficiency with human judgment, AI systems can achieve a level of accuracy, adaptability, and trustworthiness that would be unattainable through automation alone.

When presented with new situations, AI systems can automatically learn and adapt to provide better responses by drawing on pre-installed algorithms and data-driven computing technologies. “Human resource management” (HRM) is a process that can help a company provide better service to its employees by combining human interaction with AI tools (Pereira et al., 2023). A growing number of people are interested in the ethics-AI interface as a result of growing worries about human control over AI systems and their intrinsic opacity. AI systems are already replacing human HRM decision-makers, despite our inadequate understanding of the theoretical foundations of AI integration into HRM decision-making activities (Prikshat et al., 2023); Nevertheless, the emphasis on the ethical principles and values that govern the creation and utilization of AI has grown in tandem with its widespread adoption and advancements in AI capabilities (Hermann, 2022). As part of the current movement towards integrating AI with legal and ethical frameworks, Loureiro, Guerreiro and Tussyadiah (2021) propose either new sets of agreed-upon standards or a revaluation of previous moral behaviours. They argue that organizational decision analyses of existing and future systems should practically incorporate stances on AI and ethics with HRM. Algorithms are taking over more and more HRM decision-making tasks (Duggan *et al.*, 2020; Parent-Rocheleau and Parker, 2022). HR managers have been urged to forecast employee intentions and turnover using data-driven predictive analytics (Haldorai et al., 2019). This potential can only be assessed with knowledge about the company's ethical stance and strategy, within a framework that permits evaluation of decision outcomes. To mitigate the ethical and social concerns associated with AI in HRM, HR decision-makers theoretical understanding and practical experience must be considered (Charlwood & Guenole, 2022). AI definitions are expected to change as the technology develops (Hermann, 2022). Applying AI to human resource management is distinct from using other HRM terms that are enabled by technology. Prikshat, Malik and Budhwar (2023) recommended that human resource management experts build an AI assimilation model that takes into consideration all four stages of the process, including antecedents (individuals, groups, and institutions) and outcomes (operational, relational, and transformational actions) (Prikshat et al., 2023). Companies have focused their AI spending on three areas: AI employee happiness, AI service quality, and AI overall satisfaction Nguyen and Malik (2022), However, research on the ethics of algorithms utilized by AI-driven HRM systems appears to be lacking. Concerns that AI businesses would prioritize technology and profit before morality. As Charlwood and Guenole (2022) point out, a useful framework is required to assist HRM practitioners in assessing how ethics are incorporated into algorithm-based judgments and how this affects practitioners who are thinking about or already utilizing AI-driven HRM technologies.

Thus, human expertise underpins AI system reliability, impartiality, and efficiency. People's input fills gaps created by automated processes and handles domain-specific values, cultural settings, and commercial requirements to make AI outputs precise and meaningful. Human creativity and artificial intelligence speed and strength boost AI application and make the system more reliable and accountable. Thus, future AI systems will benefit from human direction on how to solve real-world challenges.

Integrating HITL in LLM-Assisted Ontology Construction

In this section, the author considers incorporating HITL approaches into LLM-based ontology development. It starts by asserting how the HITL and LLMs synergies, especially managing to demonstrate how human insight can improve the precision, specificity, and interpretative context of the AI-derived ontologies (Joachimiak et al., 2024; Makin, 2024b). It describes how HITL can overcome issues including bias, error, and domain which LLMs might not see.

Moreover, it explores the actual implementation of merging HITL and LLM capability approaches, including iterative validation, human correction, and feedback. Through this, the section shows how the section enhances the construction of knowledge representation through the combination of HITL into LLM-assisted ontology construction (Hassanin & Moustafa, 2024). The discussion also emphasizes the need for this integration in facilitating adoption of ontology development in various and sophisticated areas.(Salama & El-Gohary, 2016; P. Zhou & El-Gohary, 2016)

Construction projects are prone to legal concerns including claims, disputes, and litigations, which may cause schedule delays, cost overruns, and harm to the parties' ability to communicate and work together (Walsh, 2017). These problems mostly arise from contracts that employ natural language (Çevikol & Aydemir, 2019; Zait & Zarour, 2018). Conflicts and disputes may arise when parties to a contract fail to clearly comprehend one another due to semantic ambiguity and vagueness (Mahfouz et al., 2018; Saseendran et al., 2020). There may be problems if contractors can't predict the risks included in the contract terms (Hamie & Abdul-Malak, 2018; Hassan et al., 2021). These concerns must be carefully examined by contract management throughout the bidding and contracting process. Current methods, however, still mostly depend on labor-intensive, error-prone manual procedures and human experience. In order to support the contract management process, it requires an automated analysis approach (Hassan & Le, 2020, 2021; Khalef & El-adaway, 2021; Rameezdeen & Rodrigo, 2014).

Automating the processing of contract language using NLP techniques is an encouraging idea. Many jobs have shown their promise, such as contract reviewing, automated compliance checking, and similar case retrieval (Jagannathan & Delhi, 2019; J. Lee et al., 2019). In comparison to several domain-trained models, LLM have shown higher reasoning abilities in recent years. Their ability to comprehend linguistic subtleties and respond to a variety of human questions is a result of their training on a large text corpus. General language activities are within their capabilities, but domain-specific language may be a challenge. It could also make factual mistakes and struggle to explain their conclusions (Zou et al., 2017). Researchers are currently attempting to resolve this difficulty. Transparency and understandability in decision-making are as crucial for language models as good performance.

The Synergy Between AI and Human Expertise

AI and Human work together in harmony providing an optimal solution as the ability of machinery subtracts from the creativity of human beings (Asmi Agarwal, 2024). It is proficient at handling big data, as well as data analysis and understanding complex and repetitive tasks well (Sivarajah et al., 2017). While they certainly provide answers to questions that are posed to them in a black and white manners, they usually do not have the context, moral compass, and flexibility needed to handle more grey area situations. This is where human expertise comes in handy because it supplies the creativity, reliability of prior experience, and ethical lead that cannot be integrated into AI. In combination, AI and human produce synergistic strategies that can work faster and more accurately than when either works alone (Jarrahi, 2018). This is especially useful in areas that require precision and a context in which the results will be used in. For example, in the medical field, the AI system may input and output images or statistics which it recognizes as correlated, while the human patient-care provider inputs clinical data and interprets these findings, accounting for the patient's history, clinical guidelines or principles, and emotions.

In legal frameworks, AI will search millions of case laws and precedents, but a lawyer uses their best expertise to form arguments and judgements. Big data solutions are used by AI for decision-making, while human-centred AI is refined for accuracy, speed, and innovation (Collenette et al., 2023). Together with people, AI develops a culture of improvement. Feedback and training improve AI models' accuracy and relevance (Zirar et al., 2023). At the same time, it helps humans make decisions by providing valuable AI processing results. It solves both entities' problems and creates synergies that can support industry-wide growth business models (Kraus et al., 2022). Problems beyond AI's capabilities when solved separately can be solved by current systems using AI and human skills.

Recent advances in AI have transformed decision-making, problem-solving, and career execution. Management decision-making is one of the main areas where integration is affecting practices (Caliskan et al., 2017). Human-AI coexistence could transform decision-making, resource management, and organizational success (McKinsey Global Institute, 2016). Managers, policymakers, and researchers must understand and apply this linkage.

The greatest of human intelligence and AI technology are combined in this combo. Humans have rationality, creativity, intuition, and enthusiasm, while AI systems can analyze data, identify patterns, and analyze data quickly (Deeba et al., 2023). With these synergistic characteristics, human-AI collaboration can transform many managements decision-making processes. However, using AI in decision-making raises several difficulties and concerns. AI integration in decision-making raises the following problems and challenges. Trust, openness, accountability, and ethical issues arise. AI systems influence managerial decisions (Bamansoor et al., 2021). Job automation, skills redundancy, risk distribution, and outbound AI benefits highlight the importance of studying how human-AI cooperation affects management decision-making (Farhana et al., 2023). AI is evolving with human life, raising problems about how it interacts with managerial situations and decisions. AI is highly regarded by professionals for its capacity to analyze large datasets and acquire object knowledge. However, management supervision and legal responsibilities remain unclear. To keep organizations using AI technology ethical, efficient, and helpful, it's important to understand how humans and AI agents may scaffold decision-making. Recent advances in machine intelligence methods and technology allow machine systems to execute a wide range of tasks and application domains as well as or better than humans (Silver et al., 2017), especially in high-risk areas like public security and health care (Liu *et al.*, 2019; Norgeot, Glicksberg and Butte, 2019). Artificial intelligent systems can provide 24/7 performance, resistance to or mitigation of personal, environmental, and transitory factors, and exceptional processing capacity with increased throughput and reduced processing times. These advances have opened up new ways to integrate high-performance machine intelligence systems into decision-making across numerous domains and applications (Jin et al., 2020).

Nevertheless, there are inherent difficulties in introducing such sophisticated AI systems to vital sectors and uses, such as aviation, public safety and security, healthcare, and others. This is particularly true when it comes to the public's faith and confidence in systems that make important decisions using these systems (Xiang et al., 2023). The public may view as reckless and unjustified the practice of entrusting crucial decision-making to intricate machine systems whose inner workings and learning processes are not well understood, like deep neural networks, which are widely employed in high-accuracy image analysis. Take machine learning methods and algorithms as an example. It's not uncommon for them to make mistakes or generate unexpected results that aren't always easy to explain, assess, or fix. Kostopoulou, Delaney and Munro (2008), placing special emphasis on the importance of rigorous verification and supervision of such systems in mission-critical domains.

An additional factor pushing for the implementation of more efficient decision-making systems is growing expense of mistakes in crucial decision-making domains, such as the primary healthcare system's diagnostic error cost, which is determined to contribute substantially to total cost of public healthcare (Graber, 2013). There have been similar tendencies in other domains where operational environment demands complicated decisions.

In contrast, clustering algorithms use machine learning for real-world decision-making. Forecasting and regression (Makridakis, Spiliotis, and Assimakopoulos, 2018; Kirichenko *et al.*, 2020), classification and unsupervised learning (Schmidhuber, 2015; Z. Wang et al., 2018), and more. As "a game with nature," several criteria and approaches were established to find the best decision (Baumann et al., 2019; Newton, 2018; Ulansky & Raza, 2021). The decision maker chooses the criteria, which can be influenced by subjective attributes like optimism and risk tolerance and objective domain-describing factors like confidence, risk tolerance, decision priority, and importance. (Bavolar & Bacikova-Sleskova, 2020).

Thus, AI and human expertise enable both sides to use their strengths to eliminate their deficiencies. AI is data-driven and provides quick resolution, scaling, and more accurate outcomes, while human input adds practicality, morality, and innovation. Additionally, it improves decision-making accuracy and trustworthiness and fosters imaginative search for diverse challenges. As organizations integrate AI into their operations, balancing automation with human analysis to build robust solutions that perform in the real world will be crucial.

1.2 Research Problem

Knowledge management systems, AI, and the semantic web all rely on ontology development as a core component. The results of this study demonstrate, however, that there are significant limitations to the use of LLMs for automating the process of extracting ideas, relations, and hierarchies from massive text corpora. A number of disadvantages are characteristic for LLMs, namely: the absence of contextual comprehension, the presence of biased results in the outputs obtained with the help of LLMs, and the inability to consider the peculiarities of the field, which indicates the low quality of the constructed ontologies.

Even though automation ensures scalability and optimization, the lack of a solid validation process exposes ontology construction to potential errors that could be inherited in the applications used. To deal with this, it is proposed that the practice known as HITL (Human-in-the-Loop) could be used, which together with the LLM automation presupposes the involvement of a human specialist. Nevertheless, studies examining the best ways to integrate HITL frameworks into LLM-assisted ontology building, such as evaluating the suggested framework's efficacy and its applicability to large-scale ontology construction across domains, are still scarce.

This research seeks to offer solutions concerning such challenges by outlining the strength and weakness of using LLMs in constructing ontologies together with the HITL methods used in constructing knowledge graphs. One of the issues might be the difficulty in updating the errors made while extracting entities and defining their relationships in the course of LLM-assisted ontology building. Based on the results of this research, ambiguity can only be cleared, accurate mapping can be achieved, and the ontology generated can only be checked for consistency or checked for relevance with human intervention. Thus, this study aims at presenting a guideline on how properly constructed, relevant, better, ontology construction approaches could be constructed.

1.3 Purpose of Research

- To explore the strengths and limitations of LLMs in ontology construction.
- To review the effective integration of LLMs into ontology construction processes to improve efficiency and scalability.
- To perform data pre-processing on OSHA accident and injury data to prepare it for analysis and ensure consistency in input for further processing.
- To examine how has the human-in-the-loop (HITL) approach been applied in ontology and knowledge graph construction.
- To review the generated ontology for inconsistencies, redundancies, and ambiguities, and improve its accuracy through human intervention.

1.4 Significance of the Study

This research has considerable implications because it answers important questions and investigates essential issues in constructing ontologies – a fundamental concept in the area of knowledge representation and semantic web technologies. This study seeks to address the research gap by examining the effectiveness of using HITL together with LLM in an attempt to combine automated AI solutions with the knowledge of professionals. This contributes not only to improving the relevance and accuracy of the methodology for building ontologies but also to guaranteeing its results to be useful and ethical in the context of the problem. Consequently, the results of the current study apply to multiple fields, including artificial intelligence, data science, healthcare, finance, and education, in which ontologies are crucial for facilitating well-structured knowledge representation and management decision processes. In addition, by examining how HITL can address the problems of LLMs including bias and error to some extent, it is expected to shed light on the enhancement of the AI system.

From a wider contextual view, this research opens up the possibility of scalable and efficient construction of ontologies using AI assisted methods that incorporate the best of human inputs. Thus, through proving usefulness of cooperative AI-human orchestration, it provides a basis for subsequent developments of the intelligent systems and enhances practical implementations of the ontology-based practices.

1.5 Research Purpose and Questions

The purpose of this study is to expand current knowledge and use of modern approaches to building an ontology by combining LLMs and HITL methods. More precisely, this research seeks to discuss the prospects and challenges of using LLMs for automating the ontological process and explain the competence and incongruity of LLMs for the ontological process with solutions for scalability and efficacy problems besides the context inaccuracy and prejudice challenges. Thus, integrating the HITL methodologies, the research goals to assess the influence of the human subject expertise in improving the reliability, accuracy and domain-appropriate applicability of the LLM-based ontology construction support.

This study also seeks to examine role that HITL frameworks play in lessening the problems associated with LLMs and guaranteeing the production of quality ontologies.

That is why the research aims to identify and describe how to effectively scale ontology construction with the help of automation while ensuring human validation. Ultimately, this work hopes to add to what is already known about AI-assisted ontology construction and provide concrete suggestions for how many domains might benefit from better information management and representation.

Research Questions

- **RQ1:** What are the strengths and limitations of large language models (LLMs) in ontology construction?
- **RQ2:** How can LLMs be effectively integrated into ontology construction to improve efficiency and scalability?
- **RQ3:** What are the key steps in pre-processing OSHA accident and injury data, and how does this impact the consistency and quality of input for ontology construction?
- **RQ4:** How has the human-in-the-loop (HITL) approach been applied in ontology and knowledge graph construction?
- **RQ5:** What methods can be used to identify and resolve inconsistencies, redundancies, and ambiguities in a generated ontology, and how can human intervention improve its accuracy?

1.6 Overview of Research Methods

This study adopts a structured and ethically grounded approach to ontology construction by utilizing a publicly available dataset rather than unstructured or web-scraped data. Specifically, the research is based on the OSHA Accident and Injury dataset, which provides structured incident records spanning over two decades. Records with the keywords "electric arc" and "burn" were filtered using a purposive sampling technique, enabling targeted examination of pertinent safety incidents.

Using Python-based tools like Pandas, Matplotlib, and Seaborn for preliminary data investigation and visualisation, the analysis employs an exploratory methodology. Textual data was prepared for semantic processing using Natural Language Processing (NLP) techniques such as text normalisation, tokenisation, and lemmatisation.

Ontology construction was facilitated through a Human-in-the-Loop (HITL) framework using GPT-4. Based on cleaned event descriptions, the model produced an initial ontology that was then improved and assessed for quality attributes like accuracy, completeness, relevance, and consistency. One of the main original contributions of this research is the final ontology, which was improved by repeated human-guided refinement. It provides a new approach to enhancing automated ontology outputs in the workplace safety domain.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Background

The most popular and extensively used framework for organizing knowledge that allows for automated interpretation, reuse, and sharing is ontologies. A number of AI uses rely on ontologies; they include contextual organization, smart information retrieval, and knowledge management. Unfortunately, because to the exponential expansion of data across different industries, the process of acquiring and enriching ontologies has become labor-intensive, costly, and time-consuming. The need for automated approaches to this problem, sometimes known as ontology learning, follows. This area has seen tremendous progress because to DL models, which can infer semantic links from diverse datasets and extract concepts from large corpora. In this, Amalki and Bouzit (2025) delves into and compiles previous studies on how ontology learning may be enhanced by using DL methods. This was accomplished by selecting and analyzing 47 research publications out of a total of 27,65 that were published between 2015 and September 2024 as part of a Systematic Mapping Study (SMS). Eight improved criteria were utilized to systematically classify the studies: year of publication, kind of contribution, design of the empirical research, data type, area of application, evaluation metrics and benchmarks, and DL methods used.

In another study Al-Turki, Hettiarachchi, Gaber, *et al.* (2024) centre on creating and testing a system that can convert building regulatory texts into structured YAML representations that are good for ACC procedures utilizing the OpenAI GPT-4o paradigm. The research encompasses three distinct experimental types: few-shot learning, progressive active learning, and fine-tuning learning. The model's performance in controlling texts with few examples is supported by the few-shot learning experiment. Training the model with specialized datasets improves its performance, leading to more accurate structural and textual predictions. Improve the model's accuracy even further with progressive active learning by repeatedly integrating expert input. This study highlights the possibility of combining LLMs with active learning to automate regulatory compliance, since it shows that the produced YAML files are far more accurate in structure and semantics. This study's techniques and findings provide the groundwork for further study and real-world applications of automated regulatory compliance.

In Olga Perera (2024) looks at the new area of Generative AI, namely Big Language Models for learning ontologies. In order to analyze evaluation methods and determine the present status of Generative AI research, they surveyed the field with an emphasis on how well it applies to ontology construction tasks. They mapped out future study paths and spoke about problems with the explain ability and interpretability of Generative AI.

A growing number of academics and healthcare professionals are focusing on healthcare services and sectors including assisted care, with demands for innovative models, cutting-edge technology, and patient-centered healthcare systems. Farghaly *et al.* (2023) The principle behind person-in-the-loop reviewing is to have the computer do some of the work that the human does. As with any design of goal-oriented human-machine interaction, AI-augmented reviewing involves deciding who has control over each job and the reviewing process overall, as well as assigning duties to humans or computers. The process of assigning duties generally starts with breaking the work down into smaller tasks (e.g., assessing based on several criteria) that are then assigned based on the relative advantages of the agents as well as ethical and trustworthy factors. Opportunities and dangers vary depending on the delegation and control mechanisms used. Therefore, acknowledging that previous research has questioned LLMs' capacity to execute certain subtasks, they investigate the viability of different delegation and control structures. When it comes to jobs like evaluating originality or contribution, think about if a human principal can supervise the automated review to make sure it follows journal editorial criteria (Drori & Teeni, 2024).

A foundational component of the Semantic Web, ontologies are widely used in many industries to express ideas and the connections between them. They are clear, shared, and formalized representations of domains of knowledge. A formal vocabulary for many ideas may be provided by ontologies, which can also manage information from various sources and encourage data interchange across applications. Farghaly *et al.* (2023) Knowledge Graphs (KGs) reflect the entities and connections unique to technical papers, while ontologies offer the structural framework for defining and organizing domain knowledge inside these KGs. Nevertheless, building KGs and ontologies has historically required substantial multidisciplinary and collaborative work, which in turn has necessitated substantial time and knowledge. An attractive option is LLMs, which automate the process of generating KGs and extracting ontologies while encoding rich domain-specific and world-wide knowledge. This allows for effective information extraction with the use of deliberately developed prompting approaches (Abolhasani & Pan, 2024).

After Open AI released ChatGPT in November 2022, the general public's fascination with generative AI reached its height. The global economy stands to gain between $2.6 and $4.4 trillion USD, according to McKinsey, who predicts that generative AI will affect every industry (Neuhaus, 2023). A version of ChatGPT called GPT-4 demonstrates "sparks of general artificial intelligence," as stated by Bubeck, in its large language model (LLM). Notable computer scientists and investors have expressed concern about the possible disastrous societal impacts of modern AI systems (such as GPT-4), writing in an open letter that these systems are human-competitive at broad tasks (Ashburner et al., 2000).

On the other hand, most ontologies provide additional functions. Both machine- and human-readable descriptions of the vocabulary's semantics are provided by these ontologies, making them a public resource. Furthermore, they may stand in for actual facts and figures. One important use case for reference ontologies is their ability to facilitate collaboration among datasets that incorporate the ontology's language into their data or metadata. Bernasconi, Ceriani and Ferilli (2024) One scenario where ontologies are more appropriate is when dealing with restricted vocabulary needs, which LLMs do not provide. Ontologies also excel in situations that need for explain ability and openness, which are not present in LLMs. The information included in LLMs is dispersed among billions of parameters and is thus inaccessible to people. On the other hand, ontologies represent this knowledge symbolically, making it easy to check for errors, incompleteness, or incorrect information. (Giglou et al., 2023)

2.2 Large Language Models (LLMs) in Ontology Construction

In this work, Saeedizade and Blomqvist (2024b) explored the possibility of using Large Language Models (LLMs) to construct OWL ontologies from given ontological specifications. The ability of LLMs to provide human modellers with OWL modelling possibilities and recommendations is the focus of this research subject. Several cutting-edge models are put to the test. Their system incorporates a range of prompting tactics, such as the Zero-shot method, Decomposed Prompting, Chain of Thoughts (CoT), and Graph of Thoughts (GoT). In addition to revealing the pros and cons of the prompting tactics, the results prove that GPT-4 is now the sole model that can make good suggestions. Evidence from their study suggests that more sophisticated LLMs can outperform human novice modellers when it comes to generating OWL ideas. Being the first to comprehensively investigate LLMs' capacity to aid ontology engineers, our research represents a groundbreaking addition to this field.

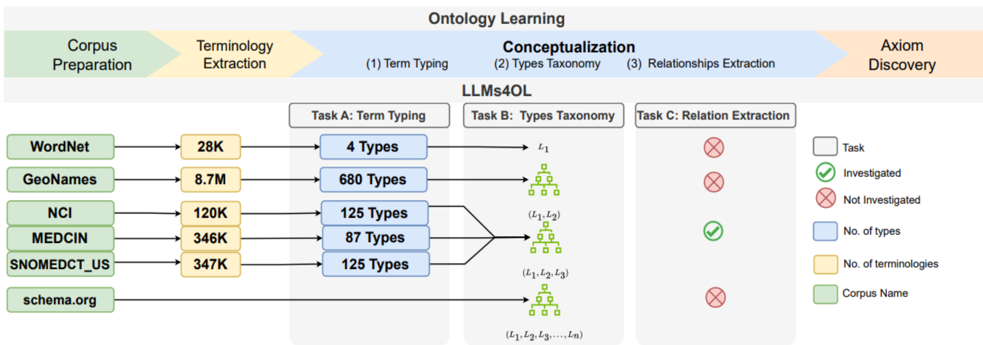

Figure 2.1: The LLMs4OL task paradigm is an end-to-end framework for ontology learning. It incorporates lexicosemantic (WordNet), geography (GeoNames), biomedicine (NCI, MEDICIN, SNOMEDCT), and online content types (schema.org). This study's experimental verification of three OL tasks is depicted by the blue arrow, which aligns with the broader LLMs4OL paradigm.

Sources: - (Gascó, 2017)

In Gascó (2017) administered a thorough assessment with the zero-shot prompting technique. Using nine separate LLM model families, they evaluate three main OL tasks: term typing, taxonomy discovery, and extraction of non-taxonomic links. In addition to lexicosemantic content in WordNet, geographical content in GeoNames, and medical content in UMLS, among many others, the evaluations incorporate a diverse array of ontological content kinds. Empirical findings demonstrate that ontology creation, which requires strong reasoning abilities and domain knowledge, is not well-suited to basic LLMs. But with proper tuning, they might be useful helpers, removing the information acquisition bottleneck during ontology creation.

2.3 Effective Integration of LLMs for Ontology Construction

Research by Joachimiak *et al.* (2024) found that AI instructors, developers, and researchers are the target audience for AIO since they are looking for a common language and set of ideas to use in the field of AI. Networks, Layers, Functions, LLMs, Preprocessing, and Bias are the six main branches of the ontology. These branches are there to help with the modular building of AI techniques, as well as to help with understanding the architectures of DL and the ethical aspects surrounding AI. The Ontology Development Kit (ODK) was used to build and maintain AIO, and AI-driven curation support allows for the dynamic updating of its content. This strategy improves new AI methodologies and approaches incorporation processes while maintaining ontology relevance to artificial intelligence growth and enhancing value for scientific researchers and education providers.

Multiple information sources in search and rescue operations create improved operational efficiency and enhanced situational awareness while speeding up decision-making processes thus increasing survival rates in incident impacts (Doumanas, Soularidis, et al., 2024). In order to integrate and reason with data from several sources, ontologies are a necessary. The development of a SAR domain ontology could be facilitated more amicably by employing an agile, collaborative, and iterative ontology engineering methodology (OEM). When it comes to completing OEM procedures, LLMs could be a game-changer. Finding out how humans and LLMs can collaborate to complete ontology engineering (OE) tasks is the main goal of this study. This work has two goals: first, to develop and evaluate an OE technique that makes use of LLMs; and second, to provide a preliminary look at using LLMs to build domain ontologies for SAR mission modelling in real-world wildfire scenarios. The main findings from this research analyze both human-robot collaborative representation of knowledge in the Search and Rescue (SAR) domain and LLM's ability to perform ontology engineering work.

The research Nie *et al.* (2024) propose a new approach for a target domain with an ontology to enhance the generation of more accurate triples from LLMs. In order to train the model to produce better triples, they mimic the way humans interpret unstructured data by using Chain-of-Thought (CoT) prompts. Building domain-specific KG is made easier and less taxing using their method, which drastically cuts down on relational expression diversity. The presented technique demonstrates the potential to lower relational expression variation through dataset experiments on TekGen. Research avenues for future study also appear in the present report.

2.4 Application of the Human-in-the-Loop (HITL) Approach in Ontology

This study concentrated on Yin *et al.* (2024) The concept of human-in-the-loop (HITL) methods is relatively new. Incorporating human experience into the model-building procedures, HITL reduces the vast amount of training data needed. The influence of generative AI tools (such ChatGPT and Mi journey) and the usage of AI to design processes have garnered more attention than the impact of different HITL methodologies on design performance. This research analyzed human-AI co-design solutions through the implementation of two independent HITL methods which included both human-learning and machine-learning approaches. Fifteen people were given both a Human-learning HITL design task and a Machine-learning HITL design challenge to complete. Four criteria were used to evaluate the participant solutions: innovation, beauty, utility, and viability. Research results demonstrated superior performance by the Human-learning HITL solution compared to the Machine-learning HITL solution specifically regarding solution aesthetics and usefulness. Instead of giving AI additional training data, the research suggests that human insights expressed via prompts are more important for advancing human-AI co-design solutions.

In Davda and V (2024) focused on the role of HITL paradigm in enhancing the quality and speed of the data annotation in AI, the proposed paradigms, and their integration. This work presents a selection of recent approaches that combines people's judgments with artificial intelligence by addressing the reinforcement of the quality of data and stability of the system using numerous applications in healthcare, self-driving cars, and language analysis. Some HITL techniques that are discussed in this work are classified, the suitability of each is assessed, and the ideas for further improvements are given.

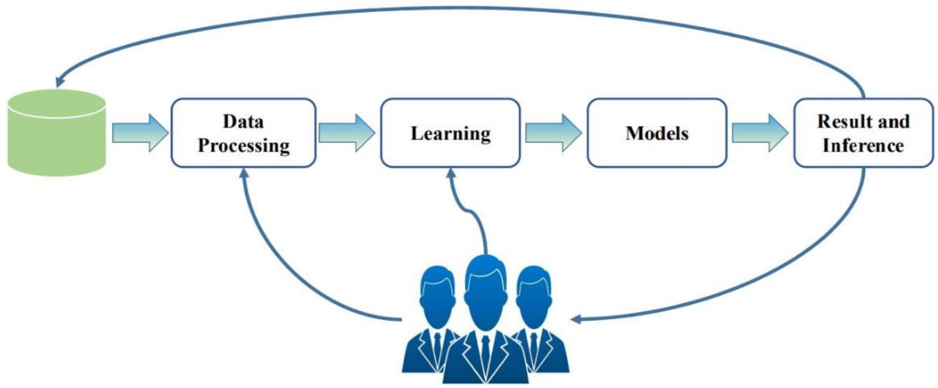

Figure 2.2: The Human-in-the-Loop (HITL)

Source: - (Boesch, 2024)

Because humans are involved in building the learnt model, interactive machine learning (IML) allows for the inclusion of human knowledge, according to a research (Estivill-Castro et al., 2022). Additionally, HITL-ML allows human specialists to direct the learning process, which may guide the learning aim towards accuracy as well as characterization and discrimination rules—the main goal of which is to separate one class from another—if necessary. Humans are also able to verify the learnt models and delve deeper into the dataset via this interaction. In order to validate, classifiers must be both transparent and easy to understand. In recent times, XAI has put an emphasis on the need of intelligible categorization for multitude of applications. Beyond parallel axis splits, they are able to create interpretable splits thanks to an IML system they implemented using parallel coordinates, which also lets them visualize decision tree classifiers. The research also shows that parallel coordinates are a good way to convey principles for discriminating and characterization. Specifically, our method is validated by the findings of the biggest usability research of an IML system, which is detailed in the paper.

2.5 Human-in-the-loop approaches for Ontology Verification and Oversight

In this study Regino and Cesar (2025) suggested using a LLM to check ontology constraints. Try again using ChatGPT-4 this time instead of humans to see if they can get the same results with ontology verification as they did with the human-in-the-loop experiment before. They find that (1) An ontology modelling qualification exam yields intermediate-to-expert results for ChatGPT-4; (2) the model has a 92.22% accuracy rate while verifying ontology restrictions; (3) raising the accuracy to 96.67% is achieved by merging model responses on the same ontology axiom expressed in distinct formalisms; and (4) compared to mistakes caused by overuse of limitations, problems connected to inadequate ontology axioms are more accurately identified. Their findings point to the possibilities of LLMs as a tool for knowledge engineering and suggest ways forward for the field's development.

In this research Drori and Teeni (2024) investigated the pros, cons, and practicality of using LLMs for academic submission evaluation while including humans. They showcase the potential, hazards, and strategies to manage them by experimenting with GPT-4 in the position of a reviewer. The reviews are organized in a way that resembles a conference review form. They serve to both assess the submissions for editorial decisions and provide writers helpful criticism based on established criteria, such as the work's contribution, soundness, and presentation. They prove that can be done by comparing and contrasting LLM evaluations with human reviews, and they find that existing AI-augmented reviewing is accurate enough to help with reviewing, but it's not perfect and it doesn't work for every scenario. They continue by outlining the benefits of AI-augmented review and posing unanswered problems. They go on to detail the dangers of AI-assisted review by bringing attention to issues including prejudice, value misalignment, and abuse. In the conclusion, they provide suggestions on how to handle these dangers.

In this research Walker *et al.* (2024) concluded that, while LLMs show promise for efficient knowledge acquisition and requirements elicitation, a broader range of skills and training is necessary for their successful implementation, especially when it comes to data comprehension and prompting. For simple quality assessment jobs, LLMs may work, but in more complicated situations, the output is difficult to regulate, and other methods of evaluation may be necessary. The purpose of this research is to document the ways in which KE stakeholders engage with LLMs, to find promising applications, and to comprehend the obstacles that prevent their efficient implementation. They come to the conclusion that copilot methods might be useful in creating procedures where generative AI aids a person or group of people.

In this research, Tsaneva *et al.* (2024) explored new methods of validation that use a HiL and a LLM-in-the-loop to enhance the precision of automated KG generation approaches. They show that by using solely LLMs, the Computer Science Knowledge Graph's automated generation pipeline might see a 12% improvement in accuracy, going from 75% to 87%. A hybrid strategy that uses both LLMs and HiL greatly improves recall and accuracy, leading to a 4% improvement in the F1 score.

In this research Zhang, A. Peñuela and Simperl (2023) discussed research themes in the field and reported on an effort to build explainable (XAI) knowledge-graph creation pipelines that include people. An interview study with thirteen members of the information engineering community and a thorough literature analysis formed the basis of their research. The review focused on frequently automated tasks in knowledge graph development and common ways to describe research methodologies and their results. They provide use cases, associated aims for XAI approaches in knowledge graph creation, and gaps in each use case in order to analyze the relevant literature. They synthesize needs for XAI method design from their practical knowledge engineering expertise, provide design blueprints, and identify research directions: (i) roles in knowledge graph building where human intervention is still necessary but where AI may be useful; (ii) enhancing the stakeholder experience via the incorporation of XAI methodologies into well-established knowledge engineering procedures; (iii) the importance of assessing the efficacy of explanations in fostering more reliable human-machine cooperation during knowledge graph creation; (iv) Customized explanations for many scenarios; and (v) testing and implementing the XAI design plan in real-world scenarios.

In this study Mateiu and Groza (2023) set out to automate the process of converting NL into DL in order to enhance ontologies. Since LLMs are the most efficient translation tools, they developed a GPT-3 model to convert NL to OWL Functional Syntax. They created NL phrase pairs with their accompanying translations for the purpose of fine-tuning. The various aspects of ontology engineering are covered by these training pairs, including instances, class sub Sumption, domain and range relations, linkages between object characteristics, complements, disjoint classes, and cardinality limitations. An ontology is enhanced in a human-supervised way using the new axioms. A Protégé plugin containing the created utility is made available to the public.

2.6 Interactive Systems: Augmenting Ontology Construction with Human Collaboration

In this research Lippolis *et al.* (2025) filled this need by investigating the use of this prompting methodology to aid in the ontology design process, with a focus on GPT-4. Studies confirm that this approach performs better than traditional and direct prompting methods which were used in previous research. Their approach, Ontogenies, uses a common collection of SPARQL-OWL queries transformed into ontology competence questions. Utilizing this approach, they were delving into various forms and degrees of knowledge refinement through MP while adhering to the extreme Design methodology, a recognized process in ontology design. Standards for evaluating ontology quality serve with ontology expert assessment as the final stage to measure both performance and quality of generated ontologies. The study contributes specific findings about metacognitive prompting and ontology design patterns to enhance discussions on ontology creation powered by LLMs.

In this research Yang *et al.* (2024) examined how existing open-source LLMs could be incorporated in the creation of KGs as (semi-)automated tools. Their process involves developing competence questions (CQs), an ontology (TBox) based on these CQs, knowledge graphs (KGs) utilizing the ontology, and finally, evaluating the KG, all with minimal to no involvement from human experts. A knowledge graph created using academic papers demonstrates the operational feasibility of their semi-automated approach to building a KG. To evaluate the quality of the responses generated by Retrieval-Augmented-Generation (RAG) and the KG ideas automatically retrieved using LLMs, they build a judge LLM that rates the output based on the ground truth. While a human-in-the-loop approach is recommended for analyzing KGs generated by machines, their findings suggest that using LLMs could simplify KG construction.

In this study Kommineni, König-Ries and Samuel (2024) investigated how open-source LLMs may be used to semi automatically generate KGs. Their process is as follows: developing competence questions (CQs), constructing an ontology (Box) from these CQs, constructing knowledge graphs (KGs) using the ontology, and lastly, evaluating the KG, all with minimal to no help from human experts. By collecting academic literature for DL approaches, the authors develop a KG to show how practical their semi-automated workflow is. A judge LLM is designed to appraise the produced material based on ground truth; it is used to assess both the replies produced by RAG and the KG ideas automatically retrieved using LLMs. Though a human-in-the-loop strategy is suggested for evaluating mechanically created KGs, their results imply that using LLMs might lessen the amount of human labour needed to build KGs.

In this study Yihang Zhao and Aryan (2024) investigated how the Onto Clean technique, together with other LLMs like GPT-3.5 and GPT-4, may be used to enhance ontologies. The two-step procedure of Onto Clean—assigning meta-properties to classes and validating a set of constraints—is crucial for determining the metaphysical quality of ontologies. The need for philosophical knowledge and the absence of agreement among ontologists make manual execution of the first step challenging in practice. The research shows that using LLMs with two different types of prompts may improve labelling accuracy. The research results point to the possibility of LLMs improving ontology refining and advise creating plugins for ontology tools to make this integration easier.

In this study, Yeh, Moritz and Hohman (2024) presented an application called Amplio which could potentially help practitioners find “the unknown unknowns” within text data and increase the data diversity through the proper identification of uncharted territories. Improving data using concepts, interpolation, and a large language model are the three human-in-the-loop methods included in Amplio. By conducting user research with 18 experienced red teamers, they show that their augmentation approaches are useful for producing model safety alerts that are diversified, high-quality, and relevant. As an example of the revolutionary power of interactive augmentation procedures, they discover that Amplio allowed red teamers to enhance data rapidly and imaginatively.

In this research Li and Klinger (2024) The GVV team created iPrOp as a novel Interactive Prompt Optimization technology to connect the automated optimization of prompts with human engagement in prompt engineering work. The optimization lobop in iPrOp allows users to evaluate developing prompts with a degree of flexibility thanks to human involvement. Using a fraction of the training data, they provide users with performance metrics, huge language model forecasts with explanations, chosen cases, and prompt modifications. Users are given the opportunity to customize the delivered prompts according to their own tastes and requirements using this technique. This approach allows non-technical domain specialists to analyze the inherent characteristics that affect the performance of prompt optimization and also helps them generate optimum prompts for their individual jobs or domains. The results of their assessment demonstrate that their approach may increase work performance by generating better prompts.

In this study Selma Wanna Fabian Parra and Pryor (2024) examined the adaptability of these models to domain-specific task planning and difficult tasks using few-shot prompting. On top of that, they provide both qualitative and quantitative assessments of prompt robustness. Lastly, they incorporate a human-in-the-loop technique to guarantee safe and interpretable job planning and execution, addressing the challenges of bringing EAI systems to actual, industrial domains. They show that a human operator and an EAI agent may work together in close proximity to one another to complete a range of inspection tasks using an AR headset to mediate the flow of information. New contributions to Embodied AI research, as far as they are aware, include applying EAI to industrial applications and using an AR headset for multimodal grounding.

In a study Al-Turki, Hettiarachchi, Medhat Gaber, *et al.* (2024) focused on creating and testing a system that converts building regulatory texts into structured YAML files that are good for ACC procedures using the OpenAI GPT-4o architecture. The research consists of three experimental paradigms including learning with few shots and learning with fine-tuning and learning with progressive active learning. Initial research indicates that the model implements successful interpretation of regulatory language through performing with limited data examples. Putting a specialist dataset through model training leads to performance enhancements through fine-tuning so the model achieves both higher textual accuracy as well as structural accuracy. The accuracy of the model is further refined by progressive active learning, which repeatedly incorporates expert input. This study highlights the possibility of combining LLMs with active learning to automate regulatory compliance, since it shows that the produced YAML files are far more accurate in structure and semantics. Here, they provide a thorough foundation for future research and practical applications in automated regulatory compliance based on our techniques and outcomes.

In this study Fernandez *et al.* (2024) has an emphasis on using spoken natural language interaction to improve operator support jobs in the manufacturing business. For the purpose of dynamically integrating this expertise into a domain knowledge graph (KG), an incremental learning technique called HIL is suggested. In-context learning for LLMs may be used to enhance other system capabilities. Experiments conducted in a real-world industrial setting with a 25% increase to the graph size show that the performance of the conversation system improves with each incremental improvement to the KG.

In this research Ciatto *et al.* (2024) proposed a novel approach to automatically populating ontologies with domain-specific knowledge utilizing LLMs as oracles, which does not depend on any specific domain Method. Starting from (i) a preliminary model consisting of interdependent attributes and classes and (ii) their function repeatedly asks the LLM to get a series of query templates, and then uses the LLM's responses to create instances of classes and properties. This means that the ontology is automatically annotated with information relevant to the domain that follows the original structure. The outcome is an automatically expanded ontology with many examples that experts may review and decide whether to retain, modify, remove, or add to based on their specific knowledge and requirements. Contribution. They generalize their method's formalization and then apply it to a case study and many LLMs. They detail investigations with their origins in the field of nutrition that automatically instantiate an ontology of food meals and their components from beginning, beginning with a classification of meals and their connections. In that section, they evaluate the produced ontologies and compare those that were obtained by using various LLMs. By lowering incorrect entities and relations by as much as 10 times, their method obtains a quality measure that is up to five times greater than the state-of-the-art, according to experiments.

2.7 Fully Automated Ontology Construction with Minimal Human Intervention

In this study Huettemann and Mueller (2025) verified that LLMs were a useful tool for improving the process of creating domain ontologies. The effectiveness of cutting-edge pre-trained LLMs was evaluated using two tasks: synonym identification and parent-child connection identification. When tested on two separate challenges, the models achieved 98% and 75.4% accuracy, respectively, in the task of automating synonym detection and relationship classification. They show that LLMs might be useful in developing and maintaining ontologies, and they also provide a methodological foundation for expanding and refining these findings. As a result, you can end up saving some time and energy.

In this research, Val-Calvo *et al.* (2025) utilized LLMs to aid in the ontology extraction process from datasets, resulting in a higher level of automation for ontology-based Knowledge Graph development. Consequently, they came up with a methodical procedure that uses LLM to increase entity quality, ontology design and construction, and data pre-processing, all of which contribute to better ontology engineering. While this work primarily focusses on ontologies, their technique may also produce mappings and RDF data. The pipeline is now a part of the Onto Genix application. The research presents the outcomes of using Onto Genix on six datasets pertaining to business operations. Despite human-developed ontologies better reflecting the most complicated scenarios, the results show that generated ontologies have patterns of coherent modelling and characteristics that are similar to those of human-created ontologies.

In this research, Baldazzi *et al.* (2024) academic and industry researchers in a wide variety of data-intensive fields are presently focusing on state-of-the-art LLMs that incorporate logic-oriented (EKGs and the larger realm of KRR methodologies. True, this kind of cooperation is critical, as EKGs provide structured insights, while LLMs provide a level of flexibility and human-centered understanding. The opposite is true with ontological reasoning, which aims to accurately handle complex tasks within a specific domain. This method addresses the core issue of LLMs' intrinsic opaqueness and encourages responsibility and confidence in AI applications by making the process more transparent and offering an explanation of the results based on their provenance. This study improves upon existing Large Language Models by introducing a novel provenance-based neuro-symbolic framework that permits ontology reasoning. In order for them to have natural language conversations with ECGs, we need to make them more domain aware and explainable.

In this research Wang, Karigiannis and Gao (2024) introduced a technique that modifies GPT-3.5 using domain-specific information for smart aircraft upkeep. In particular, GPT-3.5 is fine-tuned by investigating aircraft ontology for the purpose of curating maintenance records with an encoded component hierarchical structure. The experimental data reveals that this proposed approach demonstrates superior performance than both GPT-3.5 and GPT-4.0 in their tasks of accurate component irregularity detection and delivering reliable maintenance action suggestions. There may be other areas of manufacturing and beyond where this approach might be useful.

In this research, Tupayachi *et al.* (2024) researched the possibility of using the existing LLMs for the purpose of developing knowledge representations to bolster operations research. A workflow is developed to automate the process of creating scenario-based ontologies from existing research articles and technical manuals of urban datasets and simulations. The reasoning engine used is the ChatGPT-4 API. Prompt tweaking based on methodology, natural language processing, and the GPT are all components of this procedure. Standard protocols linked to formatting guidelines enable ontology usage to produce knowledge graphs that help generate data-based choices across multiple activities. To measure how well their process works, they compared their AI-generated ontology to the pizza ontology, which is used in tutorials for popular ontology software and is well acknowledged. The authors wrap up their discussion with a practical example of how to optimize the intricate system of freight transportation using many modes. By directing the creation of decision support strategies and crucial software components, boosting data integration and simulation coupling, and strengthening data and metadata modelling, their method promotes urban decision support systems.

In this research Reales, Manrique and Grévisse (2024) investigated the possibility of using LLMs for the purpose of extracting fundamental ideas from instructional materials. Using LLMs and ontologies like DBpedia, they suggest three distinct pathways for constructing knowledge graphs from lecture transcripts. In order to find the most important ideas (nodes) in the lesson plans, these knowledge graphs are used. When directed by ontologies, LLM-constructed knowledge graphs attain state-of-the-art performance in identifying fundamental concepts, according to the results.

In this study Lo *et al.* (2024) filled a need in the literature by presenting OLLM, a generic and scalable approach to creating an ontology's taxonomic foundation from ground up. They fine-tune an LLM using a bespoke regularize that minimizes overfitting on high-frequency concepts, rather than focusing on subtasks like particular interactions between entities. This allows them to model whole subcomponents of the target ontology. By gauging the produced ontology's structural and semantic closeness to the real world, they provide a new set of criteria for assessing its quality. To create more robust distance measurements across graphs, their metrics leverage DL approaches, in contrast to normal metrics. They found that OLLM produced more structurally sound and semantically correct ontologies than subtask composition approaches, according to both the quantitative and qualitative data they collected from Wikipedia. In addition, they show that with a little amount of training samples, their methodology might be successfully applied to new domains such as arXiv.

In this study Song and Yoon (2024) suggested a BPS system that is based on GPT, which would improve the value and efficiency of simulations by combining GPT with powerful data analytics and simulation engines. In order to provide thorough, trustworthy, and illuminating BPS settings, the ontology for GPT-based BPS is also being refined. Using this as a starting point, researchers at a high-rise residential building used CONTAM to model a multizone airflow network based on GPT. They show that GPT can get simulation data, use data mining to visualize findings, answer questions based on building knowledge, verify that design rules are followed, and suggest alternative designs. With a focus on using rigorously organized BPS engines, this research concludes that expert interventions including ontological engineering informatics are crucial.

In this study, Mandal and Connor (2024) shown that open-source LLMs like Llama-2 and Llama-3 could properly extract facts from MSTs that were tailored to certain domains. Combining in-context learning with ontology-guided triplet extraction is their method of choice. Their performance is on par with earlier approaches that used fine-tuning techniques like SPERT and REBEL, yet they only used 20 semantically similar samples with the Llama-3-70B-Instruct model. This suggests that fact extraction unique to a domain might be achieved using inference alone with very little tagged data. With this, new avenues for efficient and effective semiautomated knowledge graph generation using domain-specific data may be explored.

In this study, Hofer *et al.* (2024) highlighted the key needs for future KG building pipelines and covered the primary graph models for KGs. Afterwards, they go into the fundamentals of creating high-quality KGs, touching on subjects like quality assurance, metadata management, and ontology creation, among others. Next, they assess the current level of knowledge in KG building in relation to the needs for known popular KGs and some new techniques and tools for KG creation. In the end, they pinpoint what needs further investigation and development.

In this research Giglou, D'Souza and Auer (2023) put up the LLMs4OL method, which employs LLMs for OL (Ontology Learning). LLMs have shown themselves to be very effective in natural language processing, showcasing their capacity to grasp intricate linguistic patterns across several fields of study. The hypothesis that their LLMs4OL paradigm seeks to test is this: {Could LLMs efficiently extract and organize information from natural language text using their language pattern capture capacity in OL?} They use the zero-shot prompting approach to do a thorough examination in order to evaluate this notion. Utilizing nine separate LLM model families, they evaluate three main OL tasks: term typing, taxonomy discovery, and extraction of non-taxonomic links. A broad variety of ontological domains are also included in the examinations, including lexicosemantic knowledge in WordNet, geographical knowledge in GeoNames, and medical knowledge in UMLS, among others.

2.8 Challenges in Automating Ontology Development Using LLMs

In this study Hu *et al.* (2024) an innovative framework called LLM-Duo is proposed to enhance the automation of information retrieval from scientific papers. This means it is a hybrid of a dual-agent system and a POP method. Through the use of a prioritized BFS across a preset ontology, the POP algorithm automatically guides LLMs to acquire new information by generating structured prompt templates and action orders.

Furthermore, our LLM-Duo makes use of a specialized explorer and an assessor, who are both LLM agents. Annotation and discovery are both made more trustworthy by these two agents' cooperative and antagonistic efforts. The results show that our technique achieves better results than advanced baselines, which allows for more precise and comprehensive annotations. In a case study of speech-language intervention discovery, they use our technique to evaluate its efficacy in real-world circumstances. The speech-language therapy domain has 64,177 research publications, and our technique finds 2,421 treatments among them. This information might be very useful to the speech-language therapy community; therefore, they are compiling it into an intervention knowledge base and making it publicly available.

According to this authored Amini *et al.* (2024) the Semantic Web's ontology alignment process has long relied on comparing attributes and class labels to discover "simple" one-to-one links, which are essential for discovering linkages across ontologies. Because automating the more useful investigation of more complicated alignments is a challenging challenge, very little research has gone into this area; in actual application, ontology and domain specialists often undertake this work by hand. The recent explosion in NLP capabilities, fuelled by advances in LLMs, opens up new possibilities for improving ontology engineering processes, such as ontology alignment tasks. In order to tackle the complex issue of ontology alignment, this research explores the usage of LLM technologies. Our work is a significant advancement in automating the complex alignment job; it uses a prompt-based technique and combines so-called modules with extensive ontology content.

2.9 Scalability Challenges and Solutions in Large-Scale Ontology Construction

The study conducted by Thompson *et al.* (2023) advancement has been accompanied with an insatiable need for processing bandwidth. Computing power gains are crucial to advancements in many different fields, as this article details. Going ahead, this dependence shows that the present path is quickly becoming ecologically, technically, and commercially unsustainable. Therefore, either modifying DL or switching to alternative machine learning techniques will provide far more computationally efficient approaches, which are necessary for further advancements in these applications.

According to this study Shimizu, Hammar and Hitzler (2023) Sometimes it's not easy to modify or reuse ontologies for different uses. Their findings support this conclusion for many reasons, according to our observations: (i) disparities between the level of detail in ontologies and application cases, (ii) unclarified ontologies that might be reused, (iii) disregard for and struggle with following sound modelling practices, and (iv) not enough process support and a focus on reuse in ontology engineering tools. The Modular Ontology Modelling (MOMo) approach and its enabling tools architecture, CoModIDE (the Comprehensive Modular Ontology IDE - "commodity"), were created to tackle specific issues. Momo expands upon the well-established extreme Design approach by highlighting design pattern reuse and modular development. However, it significantly enhances this methodology by heavily using graphical schema diagrams and the accompanying technology to extract expert knowledge. Several helpful tools for implementing the MOMo process are detailed in this paper. Specifically, they assess CoModIDE's performance as a tool for graphical modelling inside the Momo approach in a comprehensive and critical manner. Commodore makes such a paradigm far more approachable and user-friendly, according to one research.

In this study Liu *et al.* (2020) determining what internet consumers may be interested in is crucial for search and recommendation engines. A web-scale ontology comprising things, ideas, events, subjects, and categories greatly aided these services. They claim that current taxonomies and knowledge bases do not uncover fine-grained ideas, events, and subjects in the online population's language style, even if they include a great deal of entities and categories. There is also no ontology that is rationally arranged among these concepts. This research introduces GIANT, a method for building a user-centric, web-scale, structured ontology. The ontology will include a lot of natural language terms that match human attention at different levels of granularity, and it will be mined from a lot of online publications and search click graphs. In order to keep the ontology organized, several kinds of edges are built. The research details the graph-neural-network-based approaches employed by GIANT and assesses the suggested methodologies in comparison to many benchmarks. Over a billion people have been impacted by Tencent apps that use the Attention Ontology, which was developed by GIANT. Attention Ontology greatly enhances click-through rates in news suggestions, according to online A/B testing conducted on Tencent QQ Browser.

According to Tudorache (2020), Biomedicine, economics, engineering, law, and cultural heritage are just a few of the many areas that have embraced ontologies. Adoption of various ontology-related standards, invention or expansion of ontology-building tools, and broader acknowledgement of the significance of standardized vocabularies and formalized semantics have all contributed to the growth of the ontology engineering profession. Methods and techniques developed via ontology engineering research are finding increasing use in production environments. There have been numerous improvements, but ontology engineering is still not easy and has many unanswered questions. This paper covers some of the open questions and potential avenues for future research in ontology engineering, as well as providing an outline of the field's evolution over the last decade.

In this research Shimizu, Hirt and Hitzler (2019) Pattern-based, modular ontologies are well-suited to FAIR data practices for a number of reasons, the most important of which are their reusability and interoperability. The development of such ontologies, however, comes with a hefty price tag; for example, knowing that a pattern exists is a prerequisite for reusing it. To get around these problems, they built MODL, a toolkit for designing modular ontologies. MODL is a library of ontology design patterns derived from many different fields and well documented. They introduce MODL as a tool, go over its applications, and provide some instances of its content in this research.

2.10 Ethical Considerations and Bias in LLM-Generated Ontologies

In this study Doumanas, Bouchouras, *et al.* (2024) explores the multifaceted realms of human and machine collaborative ontology engineering (OE). The goal of the presented work is to explore the potential of LLMs to speed up and automate the processes of collaborative OE, experimenting with different levels of LLM involvement. The proposed approach is based on a human-centered approach, that is, the HCOME approach to collaborative OE, and follows a process of exploring the declining involvement of humans and the parallel increase of LLM involvement, concluding at a level of automation where the OE is exclusively performed by LLMs. This experimentation is organized based on a series of human/LLM collaboration levels (a spectrum of OE), each one aligned to a specific OE methodology, that is, Level-0 HCOME (Human), Level-1 X-HCOME (Human and LLMs), Level-2 SimX-HCOME (LLMs and Human), and Level-3 Sim-HCOME (LLMs). The evaluation of these methodologies (one per level) is performed by measuring the similarity of the generated ontologies against “reference” ontologies (precision, recall, and F1-score of reference-to-LLM generated ontological mappings). The results presented in this study demonstrate that while LLMs significantly expedite the OE process, the accuracy and completeness of the resulting ontologies are notably enhanced by maintaining a high level of human involvement. With any luck, this research will shed light on the ever-changing dynamics of LLM-based/enhanced OE, which should lead to improvements in collaborative OE frameworks in the future.

In this study presents Wrick Talukdar and Anjanava Biswas (2023) a unique plan for textual model contextual grounding, focussing on the Context Representation phase in particular. The technique aims to enhance the models' reliability and ethical alignment through the use of a comprehensive, context-aware methodology. They lay the framework for embedding a model's behaviour in various contexts by clearly gathering and articulating important cultural, ethical, and situational aspects in a way that machines can understand. Ontologies, logic-based formalisms, and semantic web technologies are some of the tools that our method draws from in the realm of knowledge representation and reasoning. Analyses conducted on real-world textual datasets demonstrate that our methodology enhances model performance, fairness, and alignment with human expectations while maintaining high accuracy. They go on to talk about the other important parts of the framework, such as context-aware learning and interpretability/explain ability, context-aware encoding, and continuous monitoring/adaptation. This study adds to the expanding literature on responsible AI by providing a workable strategy for creating language models that are more trustworthy, dependable, and morally compatible. Findings from this research have important consequences for using LLMs in context-sensitive fields including healthcare, the law, and social services.

2.11 Enhancing Ontology Accuracy Through Human Intervention and Machine Learning Feedback Loops.

According to Caspari-Sadeghi (2023) Intelligent assessment is a covert, all-encompassing system that uses smart approaches to diagnose pupils' current cognitive level, watch their dynamic evolution, predict their achievement, and continuously update their profile. It's the basis of any AI-based educational system. Machine learning, smart sensors, educational data mining, wearable technology, and learning analytics are some of the technologies used. Adaptive, Personalized, and individualized learning and teaching might be the pivotal point of Precision Education (PE). This study explores (a) the uses of ML in sophisticated evaluation, and (b) 'Knowledge tracing and student modelling' uses DL models. Finally, the paper offers some recommendations for enhancing educational decision-making via the utilization of data and ML, and it delves into the challenges associated with using cutting-edge ML methodologies.

According to research by Confalonieri and Guizzardi (2023) explainable AI is all about creating systems that are centred on humans and can provide explanations that people can understand. Reference modelling, common-sense reasoning, and knowledge refinement and complexity management are the three key areas evaluated in the research as areas where ontologies might make a substantial contribution. Using these three axes of analysis, it summarizes and ranks many methods already found in the literature. At the end of the research, the authors go over the remaining obstacles to evaluating the efficacy and human-understandability of ontology-based methods to explanation.

In Memariani *et al.* (2021) Reference ontologies serve as a common language and database for information within a certain field. Because they build everything by hand, they can keep the quality high and are well-liked by everyone in their town. On the other hand, big domains are too big for the manual development process to handle. The ChEBI ontology is a well-known reference for the field of life sciences chemistry, and they apply a novel approach for automated ontology expansion to it. Using the ChEBI ontology's leaf node structures and the classes to which they apply, they built a DL model based on Transformers. Soon after, the model can detect and categorize chemical structures that were previously unknown to it. Compared to our earlier findings on the same dataset, the suggested model improved by 6 percentage points, achieving an overall F1 score of 0.80.

They also show how graphically representing the model's attention weights sheds light on the decision-making process, which in turn helps to explain the outcomes.

According to Michie *et al.* (2017) Addressing the threats to human health and promoting the implementation of research results in health policy and practice both need a shift in behavioural patterns. The study's authors suggest ways to put the mountain of knowledge from assessments of BCIs to greater use and encourage their widespread use. Due to the sheer volume and complexity of the evidence, more computer resources were needed to synthesize and evaluate it, as well as to make the evidence more accessible and timelier. Using AI and ML, the Human Behaviour-Change Project (HBCP) accomplished the following: (i) created and tested a "Knowledge System" that can automatically mine BCI evaluation reports for insights into behaviour change and better predict how effective interventions will be and (ii) facilitate users' ability to rapidly and effectively query the system in order to get answers to variations of the question 'What works, compared with what, how well, with what exposure, with what behaviours (for how long), for whom, in what contexts and why?' This includes practitioners, policymakers, and researchers. The HBCP was: a) provide an ontology for BCI tests and their results that connects the amount of the influence on certain target behaviours to the nature and delivery of the intervention, as well as its method of action, taking into account the moderating effects of exposure, populations, and contexts; b) create and instruct a machine to automatically extract features from BCI assessment reports using this ontology for annotation purposes; c) construct and instruct algorithms for ML and reasoning to use the annotated BCI assessment reports in order to forecast the magnitude of effects for certain permutations of behaviours, interventions, populations, and environments; d) design and develop machine and user interfaces for querying and updating the database; and e) assess the aforementioned with regard to efficiency and practicality. The goal of the HBCP is to provide users with evidence-based interventions for behaviour modification that are both current and contextually relevant. This was done to make that evidence more helpful and to back its implementation.

2.12 Ontology Quality Assessment: Identifying and Resolving Inconsistencies, Redundancies, and Ambiguities

The study conducted by Fahad Mustafa (2025) discusses the use of DL techniques especially, artificial neural networks in automating semantic analysis and the development of ontologies. The DL process helps us convert large amounts of unorganized text into valuable knowledge systems. The analysis explains how DL systems like RNNs CNNs and transformer models help find entities relations and concepts even from complex knowledge areas to build automatic ontologies. This text explores DL limits for knowledge representation such as learning scale, model interpretation and minimal data problems. The study investigates possible paths for DL knowledge representation including work with multiple data modes and creation of self-improving systems. These findings show that DL makes it possible to simplify and automate knowledge management work which can produce information systems that operate intuitively.

The focus of the study performed by Fathallah *et al.* (2024) The wine ontology is a domain-specific case study that serves as an example of a quick pipeline designed for domain-agnostic modelling. Utilizing the generated pipeline, NeOn-GPT—a procedure for automated ontology modeling—and its proof-of-concept implementation are built atop the metatheory platform. By combining the systematic approach of the Neon methodology with the generative capacities of LLMs, Neon-GPT improves the efficiency of the ontology creation process. The Stanford wine ontology is used as the gold standard for their thorough examinations of the suggested methodology. According to the findings, LLMs aren't competent enough to carry out the reasoning and domain knowledge-based procedural activities necessary for ontology building. All things considered, LLMs can't do continuous knowledge engineering without integrating with workflow or trajectory tools. But LLMs may cut down on the time and knowledge required in a big way.

According to McDaniel and Storey (2020) Research in fields like ML, the IoT, Robotics, and NLP has become increasingly dependent on domain ontologies, which codify the terminology used in a particular field, since they facilitate the interchange of data across disparate systems. However, in order to have meaningful conversations, it is necessary to guarantee the quality of these domain ontologies. Although several frameworks and criteria have been established for evaluating domain ontologies, it is still not easy to determine if they are suitable for possible applications. In an attempt to draw attention to previous work and shed light on key outstanding questions, this paper conducts a domain ontology evaluation. They categorize these evaluations into five separate evaluation methodologies and outline the current condition of each. They discuss the difficulties of domain ontology evaluation and provide solutions as well as directions for further study and practical use.

CHAPTER III:

METHODOLOGY

3.1 Overview of the Research Problem

Ontology construction is an essential part of essential systems and tools of the semantic web, artificial intelligence, and knowledge management systems. However, this work has shown that Large Language Models (LLMs) have shown much promise in automating this process by extracting concepts, relations, and even hierarchies from large text corpora but it have some important problems. A number of disadvantages are characteristic of LLMs, namely the absence of contextual comprehension, the presence of biased results in the outputs obtained with the help of LLMs, and the inability to consider the peculiarities of the field, which indicates the low quality of the constructed ontologies.

Even though automation ensures scalability and optimization, the lack of a solid validation process exposes ontology construction to potential errors that could be inherited in the applications used. To deal with this, it is proposed that the practice known as HITL (Human-in-the-Loop) could be used, which, together with the LLM automation, presupposes the involvement of a human specialist. Nevertheless, studies examining the best ways to integrate HITL frameworks into LLM-assisted ontology building are still scarce. This includes studies that evaluate the suggested framework's efficacy and its applicability to large-scale ontology building in various areas.

This research seeks to offer solutions to such challenges by outlining the strengths and weaknesses of using LLMs in constructing ontologies together with the HITL methods used in constructing knowledge graphs. One of the issues might be the difficulty in updating the errors made while extracting entities and defining their relationships in the course of LLM-assisted ontology building. Based on the results of this research, ambiguity can only be cleared, accurate mapping can be achieved, and the ontology generated can only be checked for consistency or checked for relevance with human intervention. Thus. this study aims at presenting a guideline on how properly constructed, relevant, better, ontology construction approaches could be constructed.

3.2 Research Purpose and Questions

The purpose of this study is to investigate how ontology-building using LLMs and HITL approaches can improve the process in terms of efficiency, scalability, and quality. In this study, we attempt to investigate the strengths and potentials of LLMs to build ontologies accurately, consistently and meaningfully for OSHA accident and injury data. It also considered how human intervention could improve the generated ontologies in terms of the accuracy of the resultant, and also with respect to other problems such as inconsistencies, redundancies and ambiguities.

1. What are the strengths and limitations of large language models (LLMs) in ontology construction?
2. How can LLMs be effectively integrated into the ontology construction process to improve efficiency and scalability?
3. What are the key steps in data pre-processing of OSHA accident and injury data, and how can this improve the consistency and quality of input for further ontology construction?
4. How has the human-in-the-loop (HITL) approach been applied in ontology and knowledge graph construction, and what benefits does it bring?
5. What methods can be used to identify and resolve inconsistencies, redundancies, and ambiguities in a generated ontology, and how can human intervention enhance the accuracy of the ontology?

3.3 Research Design

In this study, a mixed methods approach was used to study the pros and cons of using LLMs and HITL combinations in the process of creating ontologies. (Dawadi et al., 2021). This approach combines qualitative and quantitative methodologies. Quantitative evaluation is conducted through a paired t-test, comparing ontology quality metrics — accuracy, completeness, consistency, and relevance — before and after Human-in-the-Loop (HITL) refinement. In parallel, a qualitative evaluation leverages a structured expert questionnaire administered to 15 domain experts over three evaluation rounds, capturing both metric ratings via a 5-point Likert scale and qualitative open-ended feedback. This design ensures a holistic assessment of both objective performance improvements and expert-driven qualitative validation. Two main elements form the framework of the design: a methodical analysis of the literature and the implementation of code in Python simulations based on LLM models. Phased implementation of these two approaches guarantees a thorough examination of the problems pertaining to ontology construction, human participation, and LLM integration. The major parts of this study are as follows:

Systematic Literature Review:

- **Review of Related Work: Phase 1 -** Analyses research related to existing methods of building ontologies, Knowledge Graphs, and limitations of LLMs (reasoning, vocabulary, context) (Makin, 2024a).
- **HITL Frameworks: Phase 2 -** Investigate HITL paradigms in ontology construction, how to best practice and challenges.
- **Language Model (LM) Output Evaluation: In Phase 3,** check for inconsistencies, weak reasoning and vocabulary issues in LLM outputs and then point out where we need to improve (Chang et al., 2023).

- **Generalization of HITL Methodologies to Phase 4:** Examine the applicability of HITL frameworks into domains with requirements of compliance, context, and safety.

LLM Models, Code Implementation.

- **Ontologies from OSHA Accident Data:** Extract ontologies from OSHA admitted data, with an entity recognition step and a classification step using LLM with text classification and entity recognition powered by Python and libraries (e.g. Hugging Face, spacey) (Durmaz et al., 2024).
- **Human in the Loop Integration:** Integration with human experts to have human eyes on the generated ontology via Expert Evaluation (3 rounds) -using structured questionnaire, so that we can review and refine generated ontologies to gain higher accuracy and applicability (G. Li, 2017).
- **Ontology Evaluation:** ontology evaluation measures the ontology's precision, recall, and domain relevance to determine the practical applicability of the ontology in domains for instance, safety and compliance (Doukari et al., 2024).
- **Expected Outcomes:** To identify the weaknesses in the LLM-generated ontologies and suggest a HITL-driven improvement toward improved accuracy and contextual awareness.

Hypothesis Formulation

This study applies a paired sample t-test to quantitatively evaluate the impact of the Human-In-The-Loop (HITL) framework on ontology quality. The key ontology quality metrics examined are Accuracy, Completeness, Consistency, and Relevance. The hypotheses are formulated as follows:

- **Null Hypothesis (H_0):** There is no significant difference in the metric (Accuracy/Completeness/Consistency/Relevance) before and after HITL refinement.
- **Alternate Hypothesis (H_1):** There is a significant difference in the metric (Accuracy/Completeness/Consistency/Relevance) before and after HITL refinement.

Expert Panel Profile

To ensure robust qualitative evaluation, an expert panel comprising 15 domain specialists was engaged. The panel members were selected based on their experience in fields relevant to ontology development, artificial intelligence, workplace safety, and knowledge engineering. On average, panelists had over 12 years of experience in their respective domains. Their qualifications ranged from master's degrees in data science and safety engineering to doctoral degrees in knowledge representation and AI.

Expert ID	Role	Area_of_Expertise	Years of Experience	Department
E001	Developer	NLP	10	Data & Analytics Platform Operations
E002	Developer	Ontology Design	9	Identity & Access Management (IAM)
E003	Developer	Python	9	Site Reliability Engineering (SRE)
E004	Manager	Knowledge Graphs	13	FinOps (Cloud Financial Operations)
E005	Manager	Semantic Web	10	Platform Engineering
E006	Manager	AI Governance	13	Site Reliability Engineering (SRE)
E007	Senior Manager	Safety Compliance	14	Cloud Governance & Compliance
E008	Senior Manager	Enterprise IT	13	Cloud Governance & Compliance
E009	AVP	Information Systems	15	Cloud Infrastructure Management
E010	AVP	Operational Risk	17	Automation & Tooling
E011	AVP	Governance	18	Platform Engineering
E012	VP	AI Strategy	15	Platform Engineering

E013	VP	Organizational Knowledge	25	Data & Analytics Platform Operations
E014	VP	Cognitive Automation	22	FinOps (Cloud Financial Operations)
E015	VP	Data-Driven Policy	24	Cloud Security Operations (SecOps)

Integration of Findings:

Propose a comprehensive HITL framework combining the ideas from the literature review and Python simulations and use LLM automation and human expertise. The framework targets the development of ontologies in different domains, especially in the embedding of such ontologies within the safety and compliance domain.

3.4 Instrumentation

Multiple tools, along with instruments, help the research implement systematic literature review and code development and human-in-the-loop integration functions successfully. These platform tools have features that work together to gather data, create ontologies, and evaluate expert assessment.

Systematic Literature Review Tools

The review uses four academic databases including Google Scholar, IEEE Xplore, SpringerLink, and Scopus to locate papers and reports about ontology development, LLMs, HITL frameworks and their safety and compliance applications.

The reference management applications Zotero or Mendeley will function to arrange and manage all literature references discovered through the search process. The developed tool will properly reference all important studies while using the review stages to determine their appropriate categories.

Code Implementation

This study employs the OSHA Accident and Injury Data to systematically derive an ontology from unstructured workplace incident narratives. A data preprocessing stage with five sequential steps begins the methodology before the Event Description text undergoes standardization for analysis. The analysis of data with word clouds and bar plots helped create a structured prompt for GPT-4, which produced JSON-formatted results showing "employee -> fell from -> ladder" relationships. A JSON file collected 100 extraction results from automatic processes while showing various analytical plots to reveal the ontology output. The human evaluation followed the ontological refinement process to normalize entity descriptions and relationship terms and merge duplicate concepts through contextual groups and precise relabeling. A specially designed metrics assessment confirmed that the proposed approach effectively transformed raw textual information into domain-specific workplace safety knowledge for analysis purposes.

3.5 Data Collection Procedures

In this research OSHA Accident and Injury[1] dataset is collected from Kaggle. The terms "electric arc" and "burn" were used to search the U.S. OSHA injury database from 1984 to 2007, over a span of twenty-three years. To extract relevant information related to electrical injuries, a **keyword-based filtering technique** was employed using the terms **"electric arc"** and **"burn"**. These keywords were used to narrow down records in the dataset to those involving arc flash events. There were 532 arc flash events for which the voltage was either provided or a voltage range could be inferred. The voltage, the particular operation taking place, the arc starting mechanism, and other potentially helpful details were gleaned from the occurrence accounts. The injury reports were compiled by inspections from either the federal or state OSHA. Each report often had very little meaningful data as these individuals had been educated as general safety experts rather

[1] https://www.kaggle.com/datasets/ruqaiyaship/osha-accident-and-injury-data-1517

than electrical specialists. The system voltage may not have been recorded, but it was often feasible to determine the voltage class—low voltage if less than 1000 V or medium or high voltage if more than 1000 V—by looking at the description of the equipment or the job at hand.

The database does not include all arc-flash injuries in the United States; this is acknowledged. Due to the fact that occupational injury reports are only required by law in situations when three or more workers are hospitalized or a death occurs, many injuries get unreported to OSHA.

Data Pre-processing

Preparing data is making sense of data that isn't already in a structured form. Information gleaned from the actual environment is frequently noisy, contradictory, inconsistent, and flawed. In order to make previously incomprehensible data more understandable, a multi-stage procedure known as data preparation is employed (Agarwal, 2015). Before being input into the model, must be pre-processed to get the greatest performance possible (Edström, 2022). This research performs text normalization, tokenization and Lemmatization.

Text Normalization

Transforming text into a standardized and canonical form is known as text normalization. It include fixing typos, extending acronyms, fixing contractions, standardizing punctuation, capitalization, and other language variances to make sure textual material is represented consistently and coherently (Aliero et al., 2023) TN components play an important role in the preprocessing stage in the pipeline for TTS systems, transforming unprocessed text into a sequence of words that subsequent components of the system can further process. A multitude of approaches, ranging from weighted finite-state transducers to neural networks, have been put forward for TN. Zhang *et al.* (2023) In this research text normalization is performed using Lowercasing, Special Character Removal and Stop-word Removal.

- **Lowercasing:** The text is converted to lowercase to ensure that words like "Employee" and "employee" are treated identically.
- **Special Character Removal:** Non-alphabetic characters (e.g., punctuation, numbers) were removed using regular expressions.
- **Stop word Removal:** Using NLTK's built-in English stop word list, common words such as "the", "is", and "and" are filtered out.

Tokenization

Tokenization is a method for extracting useful information from data streams. This is a common way to talk about the initial stage of processing languages and getting data ready for artificial neural networks. In computer science and in natural processes, however, it denotes the process of reducing a complicated shape to its component parts. (Friedman, 2023) In this research, the text was split into individual words (tokens) for processing.

Lemmatization

The goal of stemming and lemmatization is to find a shared root by reducing the number of inflectional forms and obtaining similar word forms. Using a vocabulary and morphological analysis, lemmatization seeks to restore words to their fundamental form, the lemma, by removing inflectional ends. Lemmatization would try to return either saw or saw depending on whether the token was used as a noun or a verb, while stemming would maybe return only s when faced with the token saw. Words that are derivationally related are the most prevalent ones that root out. The different inflection forms of a lemma eliminate lemmatization (Lourdusamy & Abraham, 2018) (Khyani et al., 2021) In this research, Each word was reduced to its base form using NLTK's WordNet Lemmatize. For instance, "running" was converted to "run", and "employees" was converted to "employee".

3.6 Data Analysis

Model Description (GPT-4)

The data analysis in this study was conducted through a dual approach, combining traditional exploratory techniques using Python with advanced semantic processing using GPT-4. Initially, Exploratory Data Analysis (EDA) was performed to gain insights into the dataset's structure and characteristics. This involved the use of Python libraries such as Pandas, Matplotlib, and Seaborn to summarize the data, visualize distributions, detect patterns, and identify any anomalies or missing values. These steps were essential for understanding the basic trends in the data and preparing it for deeper semantic interpretation.

Following EDA, Natural Language Processing (NLP) techniques were applied to preprocess the textual data, specifically the event descriptions. This preprocessing included text normalization, tokenization, and lemmatization to ensure consistency and improve interpretability. The approach was informed by the study “An Interpretation of Lemmatization and Stemming in Natural Language Processing” by Divya Khyani et al. (2022), which guided the selection of appropriate NLP methods for effective text preparation.

Once the text data was cleaned and standardized, GPT-4 was utilized for ontology construction and semantic analysis. Structured prompt engineering was employed to guide GPT-4 in extracting key concepts, identifying relationships, and building a meaningful representation of the data in the form of an ontology. This allowed for a deeper understanding of the textual content, moving beyond surface-level statistics to uncover underlying patterns and themes within the event descriptions. By combining Python-based analysis with GPT-4’s advanced language understanding capabilities, the study leveraged both quantitative and qualitative strengths to produce more comprehensive analytical insights.

Model Evaluation

Model evaluation is the critical phase in LLM tasks. In this phase, researchers used various performance measures to accurately evaluate the LLM model. This study used several classification metrics, including accuracy, Completeness, Relevance and consistency to evaluate the model's performance.

Proposed Algorithm

Proposed Algorithm: Ontology Generation for Workplace Safety Analysis
Step 1: Environment Setup
• Install and set up Python simulation tools and Jupyter Notebook.
• Import required Python modules such as NumPy, Pandas, Matplotlib, Seaborn, NLTK, and the OpenAI API for GPT-4 integration.

Step 2: Data Collection
• Collect the OSHA Accident and Injury Data from Kaggle.
• Focus on the Event Description attribute containing the detailed narratives of workplace incidents.

Step 3: Data Preprocessing
• Pre-process the Event Description text to ensure smooth analysis and improved accuracy by applying text normalization (lowercasing, removal of special characters), stop word removal, tokenization, and lemmatization.

Step 4: Exploratory Data Analysis (EDA)
• Visualize the data using word clouds and bar plots to identify dominant keywords and understand the distribution of terms in the dataset.

Step 5: Ontology Extraction	
	• Design and implement a structured prompt for GPT-4 to extract key entities and relationships from the incident narratives.
	• Process a subset of records (e.g., 100 incidents) to generate ontologies in a structured JSON format, capturing entities and relationships (formatted as “Entity1 -> Relationship -> Entity2”).
	• Apply paired T Test to extract metrics before HITL

Step 6: Data Storage and Export	
	• Store the generated ontologies along with unique identifiers in a new dataset.
	• Export the aggregated results to a JSON file (e.g., ontology results. json) for further analysis.

Step 7: Ontology Visualization	
	• Visualize the ontology data through bar plots (e.g., top 20 entities and relationships) and directed graphs to elucidate the connections between entities.

Step 8: Human Review and Refinement	
	• Conduct a systematic review to refine the generated ontologies by addressing inconsistencies, redundancies, and ambiguities in the extracted entities and relationships.
	• Apply paired T-test to compare metrics after HITL
	• Standardize the text through normalization and contextual grouping, thereby enhancing clarity and precision.

Step 9: Ontology Evaluation

- Evaluate the refined ontology using metrics such as accuracy, completeness, relevance, and consistency to validate the quality and domain-specific applicability of the extracted information.

Finish!!!!

3.7 Research Design Limitations

This research design provides important insights regarding the HITL framework and LLM integration for ontology development although some vital restrictions need consideration:

1. **Dataset Limitations:**

The research depends on publicly available OSHA accident and injury data for its data sources. The available dataset might present two difficulties for ontology generation: outdated information as well as a lack of detailed specifics in its content. The collected data fails to demonstrate sufficient representation of safety occurrences from different industries which could reduce the transferability of research conclusions.

Any large dataset includes reporting practice-based biases that exist within the OSHA dataset and may affect the results produced by LLMs. Some safety incidents are recorded in official databases to a different degree than others which causes the data to become incomplete.

2. **Human-in-the-Loop (HITL) Constraints:**

- **HITL Integration:** The research relies on domain experts for HITL integration. One limitation is that experts are available and willing to join in the feedback iteration. In the absence of experts who are always there, the quality of feedback may suffer then the ontologies may get refined.
- **Expert Feedback:** Expert feedback provides great value but may bring subjectivity into the process of reviewing and polishing the generated

ontologies. Depending on the inputs provided by different experts, there can be either some or complete inconsistency in the feedback delivered.

3. **Technical Limitations:**
    - **Tool and Framework Limitations:** Many of the so-called NLP tools such as Hugging Face, spaCy and Protégé are utilized for such tasks but it has their own limitations in handling complex domain-specific terminologies and intricacy. Integration and compatibilities can also be challenging when there are a lot of tools (i.e., Python libraries and manually refined ontology) to be combined because it slows down the research process.
    - **Simulation Constraints:** Although the simulation is based on Python, it may not take into account all simulating aspects of the real world, especially when there is a gap between the abstractions in the text and the structured ontologies. The results of the simulations may not always match up with the challenges that are faced within the configurational, practical project of constructing large-scale ontologies.
4. **Generalizability Issues:**
    - **Specificity in Domain:** The focus of the research falls under the domain of safety and compliance, and in particular using OSHA data. This may or may not make directly applying the findings to other domains or industries. However, moving forward, these ontologies created in this study may not generalize well to other fields, including healthcare or finance, where the issues and terminologies will differ.
    - **Generalizability of HITL Methodologies:** However, the generalizability of HITL methodologies across a variety of domains is not certain due to the scalability of the HITL Framework. The proposed framework may require

varying degrees of human intervention, which depends on the level of the data complexity and domain, making it difficult to apply the proposed framework universally without modification.

5. **Time and Resource Constraints:**
   - **Research Timeline:** The time limit on the research may constrain the amount of time that can be spent on deep, iterative testing of a variety of circumstances. This necessitates that several cycles of HITL intervention require more time than initially planned, resulting in less appropriate evaluations and refinements of the behaviour. Nearly all tasks depend on a superior computational resource, especially LLMs and large-scale simulations, as the total size of LLMs and large-scale simulations is usually large. This drove the possibility that the scope and speed of the research could be constrained by the limits of the availability of high-performance computing resources.

3.8 Conclusion

This methodological framework ensures a structured and rigorous approach to ontology construction using LLMs, augmented with HITL validation. By integrating automated ontology extraction with human expertise, this study aims to develop a more accurate, scalable, and contextually relevant ontology-building process. The proposed approach provides a comprehensive mechanism for refining extracted ontologies, reducing inconsistencies, and improving the quality of knowledge representation. The findings from this study will contribute to best practices in combining LLMs with human expertise, ultimately enhancing the reliability of ontology construction in workplace safety analysis and other domains.

CHAPTER IV:

RESULTS

4.1 Research Question One

This section examines the strengths and limitations of LLMs in ontology construction, highlighting their automation capabilities, efficiency, reasoning challenges, dependency on expert validation, and the need for a hybrid human-AI approach.

Table 4.1: Strength and Limitations of LLMs in Ontology Construction

Authors	**Strengths of LLMs in Ontology Construction**	**Limitations of LLMs in Ontology Construction**
(Babaei Giglou et al., 2023)	• LLMs can automate term typing, taxonomy discovery, and non-taxonomic relation extraction. • Zero-shot prompting allows broad domain adaptability. • Fine-tuned LLMs reduce the manual effort in knowledge extraction.	• Foundational LLMs lack sufficient reasoning skills for ontology construction. • Domain-specific knowledge and expert validation remain necessary. • LLMs struggle with complex hierarchical relationships.
(Mateiu & Groza, 2023)	• Fine-tuned GPT-3 can translate natural language sentences into OWL Functional Syntax. • Can enrich ontologies in a human-supervised manner. • Available as a Protégé plugin for practical applications.	• Requires careful prompt engineering and dataset curation. • Struggles with complex ontological axioms beyond simple class relationships. • Dependence on human supervision for quality assurance.
(Mulayim et al., 2024)	• LLMs can help construct and query. semantic models in the building domain. • Reduces the expertise barrier for users unfamiliar with ontologies. • Improves interoperability using structured data representation.	• High learning curve for adoption in industry applications. • Ontology development requires significant refinement beyond LLM output. • Integration with existing frameworks is still a challenge.
(Saeedizade & Blomqvist, 2024a)	• GPT-4 can generate OWL suggestions comparable to novice ontology engineers. • Chain-of-Thought (CoT) and Decomposed Prompting improve accuracy. • Can assist domain experts in formalizing ontological requirements.	• Prompting techniques require refinement to balance accuracy and generalization. • Struggles with complex, domain-specific logic. • Only GPT-4 provided satisfactory results, limiting model selection.
(N. Chen et al., 2024)	• LLMs combined with DL improve compliance checking. • Reduces the manual effort required in rule interpretation. • Few-shot learning minimizes the need for large datasets.	• Fine-tuning is required for domain-specific applications. • Struggles with complex nested and conditional regulations. • Requires integration with domain ontologies for accuracy.

(H. Li et al., 2024)	• LLMs enhance compliance checking by integrating domain knowledge graphs. • Achieves 72% accuracy in construction scheme verification. • Well-designed prompts improve reasoning capability.	• Knowledge graphs need frequent updates to maintain accuracy. • LLM hallucinations can introduce errors. • High dependency on structured domain knowledge.
(Doumanas, Bouchouras, et al., 2024)	• LLMs accelerate ontology engineering processes. • Human-LLM collaboration improves accuracy over purely automated approaches. • Can generate domain knowledge representations iteratively.	• Fully autonomous ontology generation still underperforms human-driven approaches. • Requires substantial human oversight for validation. • Ontology completeness is limited without expert review.
(GARBACZ, 2024)	• LLMs can translate natural language into formal ontological theories. • Potential for automated ontology formalization.	• Struggles with logical consistency and formal reasoning. • Ontology structures often require manual correction. • Poor performance in handling existential and universal quantifiers.
(Kaverinskiy et al., 2024)	• Structured prompts enable LLMs to synthesize natural language from ontological structures. • High similarity scores (0.8193-0.9722) between generated and original sentences. • Useful for dialogue systems and knowledge retrieval.	• Stylistic differences remain between generated and original sentences. • Requires predefined ontological representations for training. • Limited by semantic variation in real-world datasets.

The above table 4.1 shows the construction of ontologies receives significant advancement through LLMs because they can perform automated taxonomy discovery and term typing and relationship extraction (Babaei Giglou et al., 2023) and GPT-3 and GPT-4 enable OWL syntax translation from natural language and ontology enhancement (Mateiu & Groza, 2023; Saeedizade & Blomqvist, 2024a).

The integration of DL capabilities together with compliance checking enhancement (N. Chen et al., 2024; H. Li et al., 2024) and improved interoperability in specific domain applications (Mulayim et al., 2024) constitutes the value-added by LLMs. Human-LLM collaboration fosters iterative ontology development, accelerating knowledge representation (Doumanas 2024). Natural language synthesis from ontological structures can be achieved through structured prompting methods which serve dialogue systems and knowledge retrieval applications (Kaverinskiy et al., 2024).

LLMs require human expert validation combined with prompt refinement because they face issues with reasoning abilities and managing complex hierarchical systems as well as maintaining logical consistency (GARBACZ, 2024). LLMs require structured domain knowledge from specified datasets while needing human oversight for quality assurance because updates to their systems must be conducted frequently.

Therefore, Large Language Models provide automation in ontology construction, speed up term typing, taxonomy discovery and compliance checking. This makes their integration with knowledge graphs as well as structured prompts more advanced in reasoning. Deep reasoning, complex hierarchies and logical consistency, however, are still out of reach from LLMs and, as such, need human validation. However, it is important to fine tune this for domain specific problem and there is integration challenges as well. The LLMs accelerate the ontology engineering but cannot substitute the expertise of humans. The approach is dependent on the combination of a hybrid AI driven automation augmented with a element of human help in terms of supervision. Key to this balance is the case for overcoming the risks of untrammeled autonomous generation.

4.2 Research Questions Two

The contribution of this section is to investigate how LLMs can be usefully integrated into ontology construction, to improve efficiency and scalability, both in terms of implementation strategies and challenges, and how human supervision can be leveraged to optimize the performance of LLMs.

Table 4.2: Effective Integration of LLMs into Ontology Construction Processes

Authors	**Approach to LLM Integration in Ontology Construction**	**Efficiency Improvements**	**Scalability Enhancements**	**Challenges and Limitations**
(Babaei Giglou et al., 2023)	LLMs used for term typing, taxonomy discovery, and relation extraction in ontology learning.	Reduces manual effort in knowledge extraction and structuring.	Fine-tuning improves domain specificity, making models more adaptable.	Struggles with high-reasoning tasks and domain-specific complexity.
(Lo et al., 2024)	Introduces OLLM, an LLM-based framework for automatic ontology generation.	Uses fine-tuning with custom regularization to improve model generalization.	Adapts well to new domains with minimal training data.	Overfitting to high-frequency concepts remains an issue.
(Palagin et al., 2023)	Uses ontology-based structured prompts to improve ChatGPT's meta-learning in dialogue systems.	Enhances response accuracy and relevance in domain-specific applications.	Extensible to multiple chatbot-based AI applications.	Requires extensive prompt engineering for effectiveness.
(Olga Perera, 2024)	Evaluates Generative AI methods for ontology learning, combining DL and NLP.	Automates ontology extraction from large datasets, reducing human input.	DL models improve adaptation to domain-specific knowledge.	Challenges include explain ability, semantic inconsistencies, and bias in AI-generated knowledge.
(Snijder et al., 2024)	LLMs used to match and refine mappings	Automates ontology	Hybrid approach	Domain knowledge is

	between different labor market ontologies.	alignment, reducing manual mapping efforts.	combining GPT and BERT improves cross-domain adaptability.	still necessary for high accuracy in ontology mapping.
(Mulayim et al., 2024)	Applies LLMs to construct and query ontologies for smart buildings.	Reduces reliance on specialized knowledge, making ontology use more accessible.	Generalizes across building management tasks without extensive retraining.	Complexity in model construction and maintenance.
(Ivanisenko et al., 2024)	Uses a hybrid approach of LLMs and Graph Neural Networks (GNNs) for knowledge extraction.	Improves efficiency in processing biomedical literature for ontology construction.	Allows scalable extension of biomedical knowledge graphs.	Risk of AI hallucinations and false information in generated knowledge.

This table 4.2 shows that LLMs have the ability to automate knowledge mining, taxonomy finding and relation routing to relieve the burden of human participation and increase productivity (Babaei Giglou et al., 2023). Although OLLM is not at state of the art performance, it is an advanced framework for fine tuning, custom regularization, and performs well at new domains with few training examples but over fitting remains a problem (Lo et al., 2024). Structured prompts enable ChatGPT to learn about how to learn meta for domain specific dialogue systems (Palagin et al., 2023), but is very time consuming.

With hybrid models, that combine LLMs with DL or graph neural networks (GNNs), on the one hand (Ivanisenko et al., 2024; Olga Perera, 2024; Snijder et al., 2024) on the other hand (Babaei Giglou et al., 2023) provided scalable methods for the development of biomedical and labor market ontologies. In smart buildings, LLMs are especially useful in simplifying ontologies and queries both for adoption (Mulayim et al., 2024).

However, the preceding achievements are far from real solutions, as they are subject to high reasoning restrictions and semantic inconsistencies and may suffer from AI hallucinations and developed knowledge that is biased.

4.3 Research Questions Three

This section focuses on data pre-processing techniques for OSHA accident and injury data, ensuring consistency, accuracy, and readiness for analysis. It covers data cleaning, transformation, normalization, and handling of missing values.

Dataset description

The dataset used in this study comes from Kaggle and pertains to OSHA accidents and injuries. We searched the United States Occupational Safety and Health Administration's injury database for the terms "electric arc" and "burn" from April 1984 to June 2007, a span of twenty-three years. There were 532 arc flash events for which the voltage was either provided or a voltage range could be inferred. The voltage, the current action, the arc initiation mechanism, and any other relevant details were sought after by reviewing the incident reports. The injury reports were prepared by inspectors from either the federal or state OSHA. Due to their lack of electrical expertise and training as general safety professionals, the reports generally contained little valuable information. It was typically possible to determine the voltage class—low voltage if less than 1000 V or medium or high voltage if more than 1000 V—just by looking at the equipment or the job at hand, even in cases when the system voltage wasn't recorded.

It is acknowledged that the database only includes a subset of arc-flash injuries in the United States. The legislation only mandates the reporting of occupational injuries in cases when three or more employees are hospitalized or if there is a fatality, hence many injuries go unreported to OSHA. The figure 4.1 shows a data summary

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4847 entries, 0 to 4846
Data columns (total 29 columns):
 #   Column                 Non-Null Count  Dtype 
---  ------                 --------------  ----- 
 0   summary_nr             4847 non-null   int64 
 1   Event Date             4847 non-null   object
 2   Abstract Text          4847 non-null   object
 3   Event Description      4847 non-null   object
 4   Event Keywords         4847 non-null   object
 5   con_end                4847 non-null   object
 6   Construction End Use   4847 non-null   object
 7   build_stor             4847 non-null   int64 
 8   Building Stories       4847 non-null   object
 9   proj_cost              4847 non-null   object
 10  Project Cost           4847 non-null   object
 11  proj_type              4847 non-null   object
 12  Project Type           4847 non-null   object
 13  Degree of Injury       4847 non-null   object
 14  nature_of_inj          4847 non-null   int64 
 15  Nature of Injury       4845 non-null   object
 16  part_of_body           4847 non-null   int64 
 17  Part of Body           4845 non-null   object
 18  event_type             4847 non-null   int64 
 19  Event type             4845 non-null   object
 20  evn_factor             4847 non-null   int64 
 21  Environmental Factor   4840 non-null   object
 22  hum_factor             4847 non-null   int64 
 23  Human Factor           4840 non-null   object
 24  task_assigned          4847 non-null   int64 
 25  Task Assigned          4847 non-null   object
 26  hazsub                 4847 non-null   object
 27  fat_cause              4847 non-null   int64 
 28  fall_ht                4847 non-null   int64 
dtypes: int64(10), object(19)
memory usage: 1.1+ MB
None
```

Figure 4.1: Data summary

Ontology Evaluation

The refined ontology was evaluated against the automated ontology to assess the quality of relationships extracted during the initial generation. The evaluation was conducted using the following metrics:

1. **Accuracy Relationships**:

The accuracy metric determines how many automated ontology relationships successfully correspond to relationships found in the refined ontology.

- **Formula:**

$$Accuracy = \frac{Correctly\ Matched\ Relationships}{Total\ Extracted\ Relationships} \ldots \ldots (4.1)$$

2. **Completeness Relationships**:

This measurement determines how many relationships the extraction method retrieved relative to the predicted relationships in each record.

- **Formula:**

$$Completeness = \frac{Total\ Number\ of\ Extracted\ Relationships}{Total\ number\ of\ Records} \ldots \ldots (4.2)$$

3. **Relevance Relationships:**

The evaluation checks whether extracted relationships match the pre-defined parameters of domain-specific rules. The workplace safety ontology contains relevant terms which include "employee", "fall" and "injury".

- **Formula:**

$$\text{Relevance} = \frac{Relevant\ Relationships}{Total\ Extracted\ Relationships} \ldots \ldots (4.3)$$

4. **Consistency Relationships:**

This measure checks whether all extracted relationships exist without duplicates or inconsistencies throughout distinct recorded entries.

- **Formula:**

$$Consistency = \frac{Number\ of\ Records\ with\ Unique\ Relationships}{Total\ Number\ of\ Records} \ldots \ldots . (4.4)$$

Exploratory data analysis

The EDA phase analyzes the Event Description attribute in workplace incident reports to discover important patterns throughout the data. The analysis uses multiple visual methods to reveal the most common terms, entities, and actions that appear in the dataset.

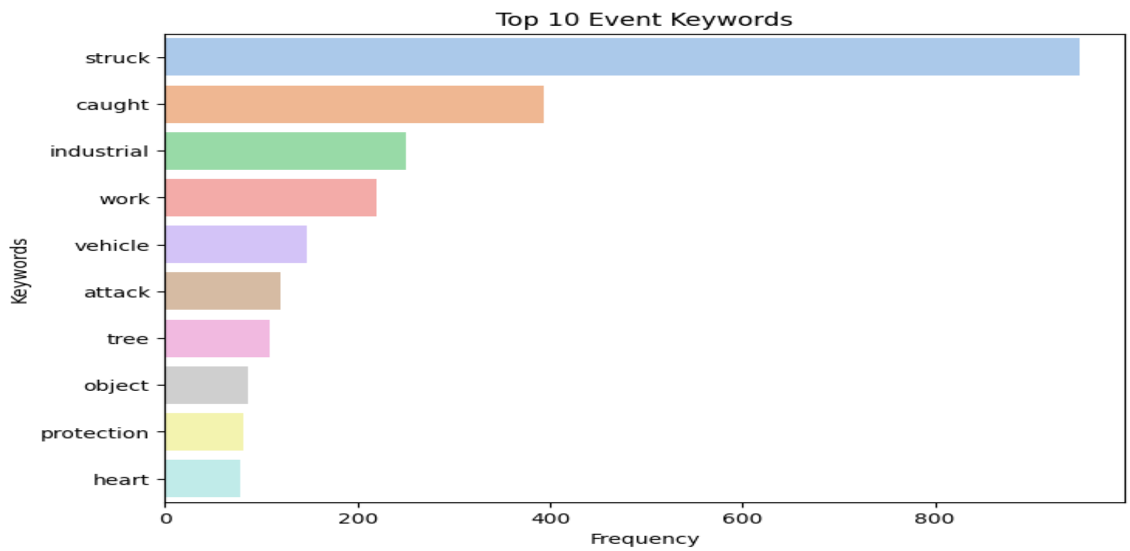

Figure 4.2: Top 10 Event Keywords.

Figure 4.2 shows the "Top 10 Event Keywords" by displaying the most common terms within the Event Description to illustrate major discussed entities along with actions. The bar chart shows keyword frequency through its x-axis alongside the y-axis that lists "struck," "caught," "industrial," "work," "vehicle," "attack," "tree," "object," "protection," and "heart" as its top 10 keywords. The word "struck" appears most frequently as the bar chart shows its peak followed by "caught," and then the other keywords show minimal occurrence. The graphical display reveals regular patterns among discussed topics that may arise from incidents, workplace hazards, or occupational circumstances.

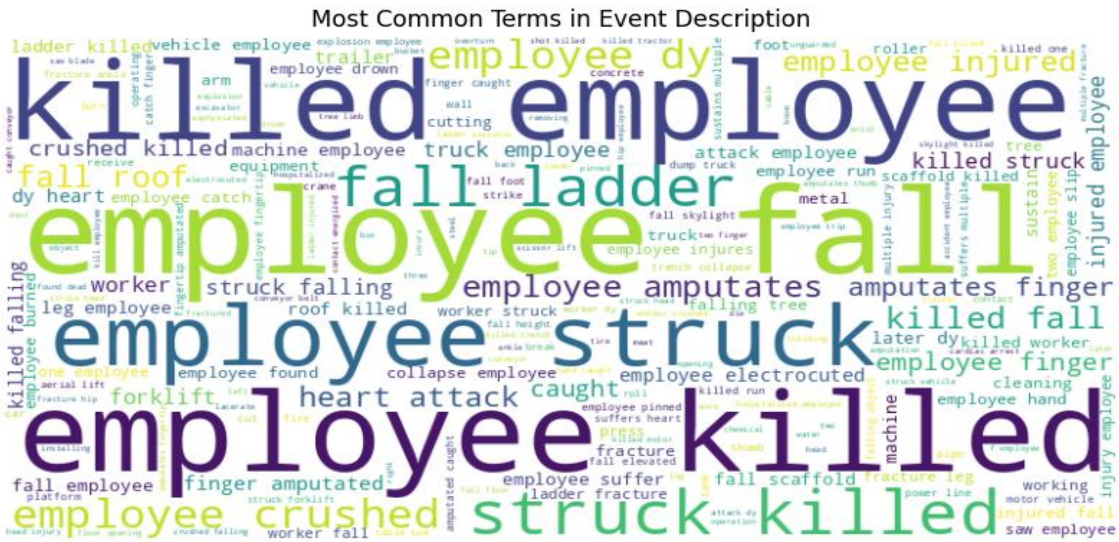

Figure 4.3: Most Common Terms in Event Description

In the above figure, 4.3 presents the word visualization of the Most Common Terms in Event Description, which resulted from dataset preprocessing. The workplace incident reports in the word cloud demonstrate three terms of highest frequency: "killed," "employee," and "fall." The dataset contains workplace-related accident and injury material as supported by the terms "struck," "crushed," "ladder," "roof," "finger," and "amputates". Words appear in different dimensions and colors between blue, green, yellow and purple, while their sizes represent term occurrence frequency in the dataset. The distributed arrangement of words throughout the visualization helps readers spot the major patterns in workplace accidents. These features allow improved analysis of workplace injury and accident causes by creating an effective visual accident.

Ontology Generation Process

Prompt Design

To ensure GPT-4 accurately extracted the required information from the **Event Description**, a structured and clear prompt was created. The prompt was designed to guide the model to focus on identifying key entities and relationships within the incident report narrative.

Prompt Template:

```
Extract entities, concepts, and relationships from the following incident
report:
{report}
Provide the result as structured JSON with the following keys:
- Entities: List of entities or concepts in the report.
- Relationships: List of relationships between entities, each as 'Entity1 -
> Relationship -> Entity2'.
```

Figure 4.4: Prompt Template.

Figure 4.4 shows entities that serve as the primary components of incidents, which include employees and ladders and machines and falls. The formal expression of entity relationships appears as "Entity1 -> Relationship -> Entity2" in structured formats. The structured framework includes relations like "employee -> fell from -> ladder" alongside "machine -> caused -> injury." The method establishes a defined framework that enables GPT-4 to perform consistent systematic data extraction from incident reports and similar datasets.

Automation

The ontology generation process ran in an automated fashion through Python to process Event Description attributes within 100 incident reports for consecutive iterations. GPT-4 received each Event Description record through a well-formatted prompt during the processing phase. The input sent to GPT-4 generated a JSON response that included Entities that displayed extracted key terms or concepts as well as Relationships that showed connections between these terms. The information extraction process for ontology generation followed a standardized framework to deliver ordered structured results.

- **Example Input and Output:**

Input (Event Description):
```
An employee fell from a ladder while performing roof repairs,
resulting in injury.
```
Output (Generated Ontology):
```
{
    "Entities": ["employee", "ladder", "roof", "injury"],
    "Relationships": [
        "employee -> fell from -> ladder",
        "ladder -> used for -> roof repairs",
        "employee -> suffered -> injury"
    ]
}
```
Figure 4.5: Example Input and Output.

The process was implemented programmatically to handle multiple records efficiently, and the results were stored in a new column called Generated Ontology.

Output Format

The ontology generation system produced standardized JSON data with summary as the unique incident report ID. The Generated Ontology organized information into two principal sections, with Entities consisting of important narrative terms and Relationships depicting semantic entity connections. The defined format produced consistent results, which made the analysis more effective.

- **Output Example**:

```
{
    "summary_nr": 123456,
    "Generated Ontology": {
        "Entities": ["employee", "fall", "ladder"],
        "Relationships": ["employee -> fell from -> ladder", "ladder ->
caused -> injury"]
    }
}
```
Figure 4.6: Example Input and Output.

This structured format ensured easy integration with subsequent analysis and visualization tools.

File Export

The subset of records underwent processing which led to ontology generation until the researchers produced JSON files for storage and future usage. The file named ontology results. JSON organized incident information by using a summary to identify occurrences while presenting the extracted Generated Ontology content. The JSON file gathered all recorded data to provide elementary support for multiple downstream processes that involved entity relationship visualizations along with workplace incident pattern analysis and ontology human inspection.

Visualization of Ontology Data

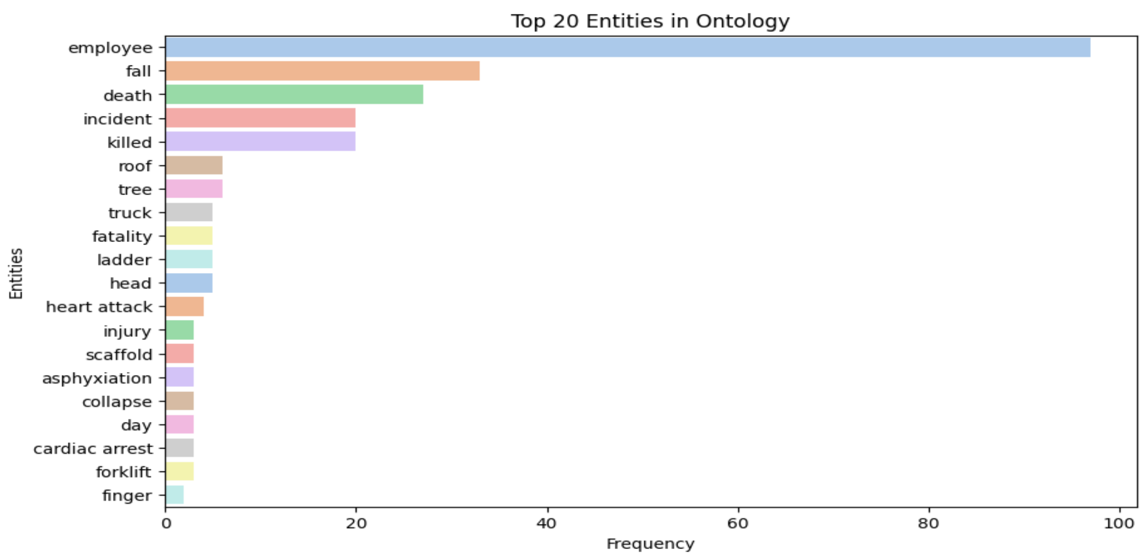

Figure 4.7: Top 20 Entities in Ontology.

Figure 4.7 shows the "Top 20 Entities in Ontology" bar plot that demonstrates the prevalence of entities within workplace safety incident reports. The y-axis shows entity types ranging from employee through fall and death up to incident thus indicating workplace injuries and fatalities are the main focus of the reports. Various accident types and hazards appear in the workplace safety incident report entities, including “killed,” “roof,” “truck,” “fatality,” “ladder,” “injury,” “scaffold,” “asphyxiation,” “collapse,” “heart attack” and “forklift.” The documentation section above the graphical representation uses two important points to analyze the data relationship between entities and workplace security elements and the human perspective of reported incidents through phrases such as “employee” and “injury.” The chart displays employee incidents as the longest coloured bar to highlight their importance in workplace safety risks through clear visual depiction.

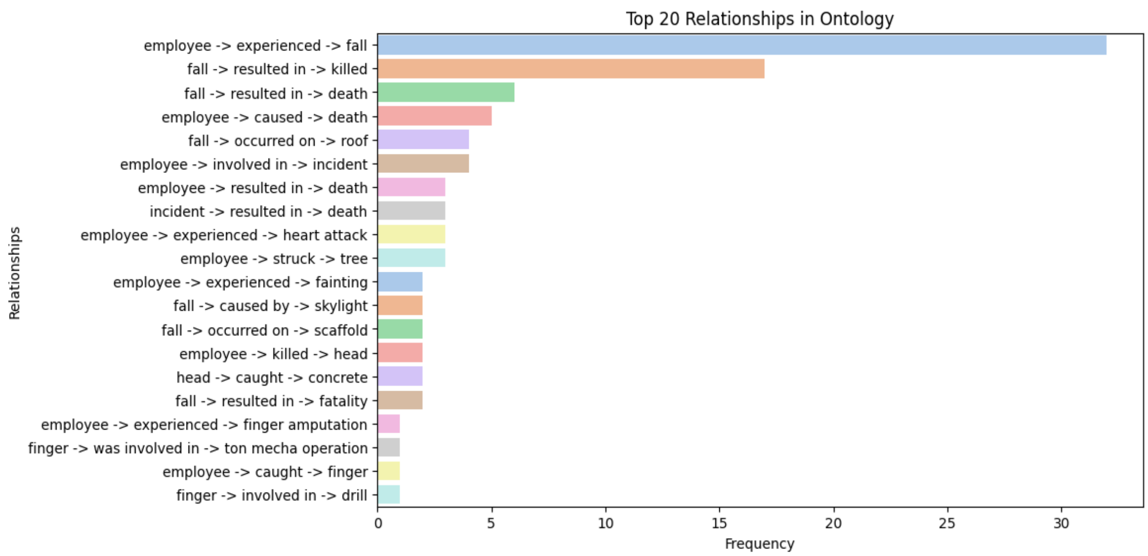

Figure 4.8: Top 20 Relationships in Ontology.

In the above figure, 4.8 shows "Top 20 Relationships in Ontology," which displays the frequency of workplace safety incident relationships. Falls represent the dominant workplace injury relationship which links employees to experience falls. Falls that result in death or fatal consequences appear twice in the data as "fall -> resulted in -> killed" and "fall -> resulted in -> death." The key relationships between falls that occurred on roofs and scaffolds establish height-related falls as particularly hazardous events. The record shows workplace danger possibilities through two separate sets of data points which show "employee -> experienced -> finger amputation" incidents while showing other events where "head -> caught -> concrete." Most of the recorded workplace interactions focus on fall incidents along with fatal outcomes and major injuries which strongly demonstrate the necessity for improved safety protocols.

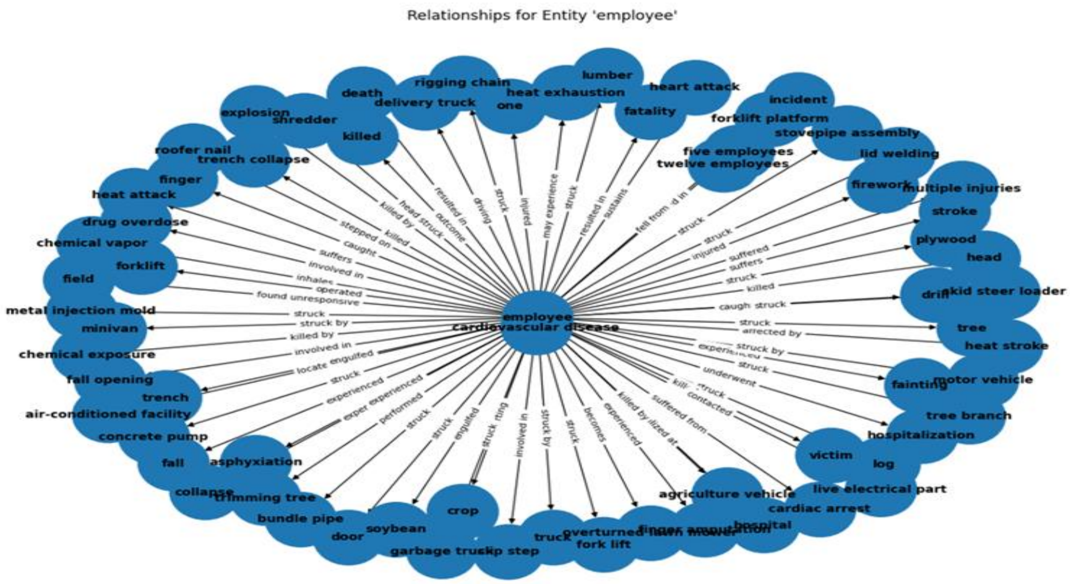

Figure 4.9: Top 20 Relationships in Ontology

Figure 4.9 shows "Top 20 Relationships in Ontology" displays the frequency of workplace safety incident relationships. The main employee-experienced fall relationship shows workplace falls remain the primary cause of injuries leading to accidents. The relationships "fall -> resulted in -> killed" and "fall -> resulted in -> death" show how falls produce fatal outcomes. The key relationships between falls that occurred on roofs and scaffolds establish height-related falls as particularly hazardous events. The record shows workplace danger possibilities through two separate sets of data points which show "employee -> experienced -> finger amputation" incidents while showing other events where "head -> caught -> concrete." These workplace relationships focus primarily on falls and deaths and severe injuries which emphasizes the urgent need for workplace safety enhancements.

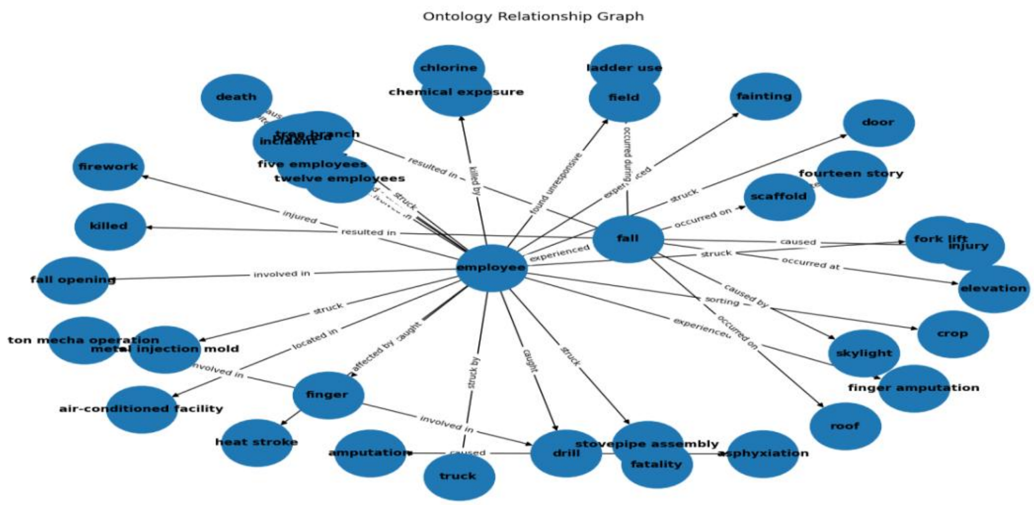

Figure 4.10: Ontology Relationship Graph.

Figure 4.10 shows a graph visualization named "Ontology Relationship Graph" to present the first 50 relationships from the ontology. The graphical representation includes nodes, which depict entities and edges that symbolize their relationships with each other. Employees occupy a central position in this node because they serve as the main subject in workplace events. The entities "fall," "injury," "death," "killed," "firework," "chemical exposure," "scaffold," and "amputation" keep directed edges connecting them to "employee" to demonstrate both incident involvement and causal links. The visual design of the graph reveals intricate workplace safety hazards because it extends entities starting from the "employee" central point to illustrate recurring safety issues. This high-level ontology structure enables risk understanding because it detects recurring hazards and their interconnected elements according to the section. The visual representation highlights recurring incidents affecting employees to generate information for creating safety programs that prevent future occurrences.

Human Review

The ontology data was reviewed to ensure consistency, correctness, and completeness of the extracted entities and relationships. The review identified several issues that could impact the quality and usability of the ontology. Below are the key findings and recommendations.

1. Issues Identified

- **Inconsistent Capitalization**

An inconsistency existed among capitalized words in extracted entities because some names were detected as "employee" against others appearing as "Employee." The varying capitalization between different parts of information can create duplicate entries or analysis misinterpretations which reduces data accuracy. It is suggested to normalize entity and relationship names into either lowercase or title case format to achieve consistency in the data and enhance its reliability.

- **Redundant or Generic Entities**

The presence of repeated entities together with generic concepts in the ontology produces a lower descriptive quality that diminishes its overall value. The ontology contained multiple instances of the basic term "employee" without extra details, while it also utilized the generic term "incident" with no specific case definition. The ontology becomes less precise and useful because of unspecified details. More detailed terminology needs to replace existing terms for enhancement. The description becomes more meaningful when "employee" type terms are replaced with designated descriptions such as "injured employee" or "supervisor." Using replacement terms "firework accident" and "machine failure" instead of "incident" both enhances precision and makes the ontology more informative.

- **Vague or Ambiguous Relationships**

The extracted data contained relationships that were either unclear or imprecise and used terms such as "experienced" and "occurred on." The phrase "fall -> occurred on -> roof" within the data fails to show what sort of situation really transpired. Using ambiguous terminology in relationships leads to reduced effectiveness during downstream operations that need reasoning or querying. It is advisable to transform imprecise data expressions into clearer versions like "fall -> occurred from -> roof" or "employee -> fell from -> roof" when improving both data clarity and usability.

- **Formatting Issues**

Two different formatting styles appeared in the relationship mapping where the first example used "Finger -> Involved In -> Drill Accident" with Title Case and the second example used "employee -> hospitalized -> incident" with Mixed Case. The varied formatting of relationships in the ontology leads to confusion for analysis tools which may result in extra duplicates that decrease its accuracy. For consistent data processing, it is advisable to use a single formatting standard for relationships such as all lowercase or all title cases.

- **Lack of Contextual Grouping**

The current organization separates fireworks and incident entities into distinct categories instead of grouping them together, which results in reduced ontology effectiveness and potential understanding inadequacies. Without contextual grouping of related entities, the ontology experiences diminished conceptual coherence because essential relationships fail to link relevant concepts properly. The enhancement of clarity along with completeness requires a combination of related entities under compound entities such as "firework accident."

2. Corrections Made

- **Entities**

All entity names throughout the ontology received normalization by transformation to lowercase letters to establish uniformity. Removal of redundant entities helped prevent duplicative data entry and enhanced understanding of the system. In relation, entities merged into single compound expressions as needed to improve contextual insight. The two terms "firework" and "incident" merged into "firework accident" improved the ontology's structured representation capabilities.

- **Relationships**

Lowercase formatting ran throughout all relationship definitions to maintain consistency in the ontology. The definition of ambiguous terms received specific language changes in order to improve clarity and precision in the modelling process. The phrase was adjusted by changing experienced to suffered and occurred on to occurred from to refine its meaning clarity. The modifications enhance both the accuracy and communicative power of the ontology system.

- **Duplicates**

The ontology standardizes every relationship entry to lowercase formatting to maintain consistency. The definition of ambiguous terms became more precise to improve the quality of representation. The term ‘experienced’ was changed to ‘suffered’ to make the message easier to understand, as was ‘occurred on’ to ‘occurred from’ to help explain the dynamics of the relationship. These enhancements improve the precision and improve representation ability of the ontology.

3. Key Improvements

Several key enhancements are involved while improving an ontology. Second, consistency is achieved through the use of uniform capitalization and formatting for entities and relationships in the structure, therefore, giving it a more standardized and readable format. Using precise terms that make the ontology more descriptive and easier to understand creates clarity. Redundancy is further reduced by removing duplicate or generic entities which increases the overall quality and efficiency. Finally, compound terms are introduced to increase contextual relevance of the ontology by offering better context for downstream tasks, which makes the ontology more meaningful and useful in practice.

4. Ontology Refinement:

- **Loading the Ontology**

The first step of ontology refinement began with the loading of the raw ontology dataset (ontology_for_review.csv) which was the produce of the ontology generation phase Entities and Relationships extraction process. Ontology refinement needed this dataset to be thoroughly examined and to be confirmed that the data was both accurate and complete.

- **Normalization**

The dataset-maintained consistency through normalization techniques, which were applied to Entities and Relationships. The key steps included:

  - The normalization process includes converting every text entry to a single lowercase format.
  - The process of removing extra spaces from text strings creates a clear representation of the data structure.

- **Refining Relationships**

A systematic examination of ontology relationships occurred to remove ambiguities while improving understanding of the information. Key refinements included:

  - The analyst switched ambiguous words ("experienced") into clear technical expressions ("suffered").

- The definition of "hospitalized" was developed into a specific statement that added "due to" as the factor causing hospitalization.
- Standardizing verbalizations throughout all interconnected relationships will improve understanding of the information.

- **Combining Related Entities**

A combined approach of entity consolidation created compound entities from entities that shared close relationships to maximize context representation. The combination of "firework" and "incident" created the single entity of "firework accident" for more effective data understanding.

5. Experimental Results

```
Evaluating Relationships (Automated vs Human-Reviewed)...

Relationship Metrics:
accuracy_relationships: 0.69
completeness_relationships: 2.46
relevance_relationships: 0.78
consistency_relationships: 1.00
```

Figure 4.11: Evaluation Metrics for Relationship Extraction

Figure 4.11 illustrates the evaluation assessment of automated ontology relationships against the refined ontology model. The automatic system achieved a 0.69 accuracy rate by matching relationships, which were also present in the refined ontology. The recorded completeness reached an average of 2.46 documented relationships per record. Domain-specific rules indicate that 0.78 of relationships established in the automated ontology proved relevant. All 1.00 of database records demonstrated the presence of non-duplicate relationships which passed the consistency metric test. The results demonstrate how well the automated ontology manages precision and how it handles fullness and suitability and maintains coherent relationship systems.

Table 4.3: *Key Performance Indicator (KPI) Comparison:*

KPI	Before HITL	After HITL
Accuracy	0.51	0.78
Completeness	1.38	2.46
Consistency	0.81	1.00
Relevance	0.61	0.78

Table 4.4: *T-Test Results:*

Metric	t-stat	p-value	Significance
Accuracy	-21.824	0.0000	Significant
Completeness	0.901	0.3693	Not Significant
Consistency	-11.084	0.0000	Significant
Relevance	-18.032	0.0000	Significant

Table 4.5: *Expert Questionnaire Summary:*

Metric	Average Expert Score (1–5)
Accuracy	4.3
Completeness	4.5
Relevance	4.4
Consistency	4.8

4.4 Research Questions Four

This section explores the application of the Human-in-the-Loop (HITL) approach in ontology and knowledge graph construction, highlighting its role in improving accuracy, resolving ambiguities, and enhancing AI-driven knowledge representation.

Table 4.6: Application of Human-in-the-Loop (HITL) in Ontology and Knowledge Graph Construction

Authors	**HITL Approach in Ontology & KG Construction**	**Key Benefits**	**Challenges & Limitations**
(Palma, 2024)	Uses HITL for knowledge graph creation in cultural heritage, guiding semi-expert users through structured data extraction and relation discovery.	Enhances user interaction in KG creation, improving narrative-based knowledge representation.	Requires domain expertise for validation; lacks automation in narrative linking.
(Hoseini et al., 2024)	HITL is integrated in semantic data lakes to validate and refine knowledge graph mappings, ensuring high-quality metadata alignment.	Improves metadata consistency and semantic interoperability in big data environments.	Scaling HITL for large datasets remains a challenge; human intervention is time-intensive.
(Carnevale et al., 2024)	Proposes HITL as a means to align ontology-based AI systems with ethical and contextual human insights.	Ensures AI-driven ontologies reflect human-centric ethical considerations.	Complex ethical trade-offs require continuous human oversight.
(Gil et al., 2020)	HITL framework helps refine CPS-related ontologies by integrating human decision-making into KG updates.	Improves adaptation of ontologies in real-time systems, enhancing dynamic updates.	High reliance on real-time human intervention for quality control.

(M. Chen & Zhang, 2020)	HITL applied to ontology learning by integrating human knowledge into ML-driven ontology refinement processes.	Improves ontology accuracy by correcting machine-generated inconsistencies.	HITL requires efficient user feedback mechanisms to scale.
(Nobani, 2022)	HITL used in labor market intelligence (LMI) for taxonomy refinement, integrating expert-driven validation in knowledge graph construction.	Enhances labor market ontologies by ensuring domain-specific accuracy and consistency.	Time-consuming and dependent on expert availability.

According to the review of previous studies, presented in above table 4.3, Human in the loop (HITL) approaches influence the construction of ontology and knowledge graph (KG) through human expertise played in refining AI generated structures for different domains. HITL increases the capability of narrative based knowledge representation in cultural heritage, however relies on domain experts for validation (Palma, 2024). It achieves high quality metadata alignment in semantic data lakes, which is still limited by the time consuming and infeasible human intervention in the process (Hoseini et al., 2024).

Similar to the ethical considerations, HITL aligns ontology-based AI with ongoing human supervision for complex trade-offs (Carnevale et al., 2024). It improves CPS ontology update with the real-time, but requires human decision making for quality control (Gil, Albert, Fons, & Pelechano, 2020. HITL is applied to machine learning driven ontology refinement that corrects inconsistencies but also requires scalable efficient feedback mechanisms (M. Chen & Zhang, 2020).

Expert driven validation provides expert accuracy but is time consuming and requires expert availability in labor market intelligence (Nobani, 2022). HITL improves ontology construction with the aid of human expertise but it relies on manual intervention making them difficult to deploy in terms of their automated, scalability, and efficiency.

4.5 Research Questions Five

This section focuses on reviewing the generated ontology to identify inconsistencies, redundancies, and ambiguities. It emphasizes the role of human intervention in refining and enhancing accuracy for improved knowledge representation.

Table 4.7: Human Intervention for Improving Ontology Accuracy

Authors	**Ontology Issues Identified**	**Human Intervention Methods**	**Accuracy Improvements**	**Challenges & Limitations**
(Kommineni et al., 2024)	Inconsistencies in automatically extracted KG concepts and relations.	HITL is used to validate LLM-generated knowledge graphs and refine ontology structures.	Improved relevance and correctness of knowledge graphs through human validation.	Requires expert oversight to validate AI-generated ontology structures.
(Keloth et al., 2018)	Missing child concepts in ontologies due to structural variations.	Human experts review and validate suggested missing concepts from other ontologies.	Algorithmic import achieved statistical significance in ontology enhancement.	Ontology importation requires human validation for contextual accuracy.
(Tsaneva & Sabou, 2024)	Incorrectly modeled concepts and controversial viewpoints.	Human computation and expert crowdsourcing	Improved trust in AI-driven knowledge systems by ensuring	Crowdsourced evaluations require effective

		applied for ontology validation.	accurate concept representation.	task design and quality control.
(Nowrozy et al., 2024)	Ambiguities in access control ontology leading to security issues.	HITL used to refine policy-based ontologies in security applications.	Enhanced policy alignment with real-world compliance requirements.	Domain experts needed to interpret and validate policy rules.
(Ben Abacha et al., 2016)	Logical inconsistencies and concept misalignment in medical ontologies.	Human experts validate automatically generated natural language questions.	Reduced manual effort in ontology validation while maintaining quality.	Complex ontologies require extensive question generation and evaluation cycles.

As can be seen in the above table 4.4, the process of ontology construction is challenging because there are inconsistencies, and missing concepts, ambiguities, and logical misalignments that need to be refined and verified by human intervention. However, HITL methods serve to validate LLM's generated knowledge graphs, but the expert oversight is needed (Kommineni et al., 2024). Missing child concepts, which are often caused by structural variations, can be reviewed and validated by human experts and do show statistical significance in ontology enhancement (Keloth et al., 2018).

Expert crowdsourcing helps in addressing controversial viewpoints and misleading modeling and it facilitates trust in AI driven knowledge systems, however the task design and quality control are critical (Tsaneva & Sabou, 2024). Human experts are used in security applications to refine access control ontologies by aligning policy with real world compliance, but expert interpretation is needed (Nowrozy et al., 2024). Human validation of the questions generated by AI reduces the manual effort, while ensuring the quality, however complex ontologies require comparatively high evaluation cycles (Ben Abacha et al., 2016).

Nevertheless, improvements in ontology accuracy can be greatly improved by human intervention but there are still challenges of expert dependency, scalability, and validation efficiency.

4.6 Summary

These findings point out the possibility of using Large Language Models (LLMs) for ontology construction by automating simple tasks such as taxonomy discovery, term typing, and relation extraction and decreasing manual labor and increase in efficiency. However, LLMs also have their share of limitations such as lack of reasoning capabilities, inability of working with hierarchical data structures, as well as the requirement of constant validation by experts to deliver output, which is accurate.

The ontology-based methodology investigates how the methods of increasing data integration and analytical services can be promoted for domain-based decision making. The research method involves four stages: a description of the datasets and an evaluation of ontology concepts, a generation and an exploratory data analysis of ontology.

Analysis of the dataset is done to ensure appropriateness of the study by checking the structure and attributes and relationships to the dataset. The evaluation of ontology is to specify the quality in which the domain knowledge is represented in the ontology. The process of exploratory data analysis (EDA) identifies patterns together with trends and clear indications of inconsistencies found in the dataset which leads to increased ontology development precision. The ontology generation process converts unprocessed data to structured knowledge representations which creates semantic understanding and interoperability networks. The research implements recognized methods combined with best practices throughout the ontology development process for achieving both precision and stability in addition to user-friendly applications.

By integrating DL and structured prompts, LLM effectiveness is improved, and in particular in the domain of applications such as compliance checking, biomedical knowledge graphs and smart building ontologies.

However, although these advancements have been made, the goal of maintaining logical consistency, semantically consistent instruction, and defeat hallucinations done by AI remains an unresolved issue.

Furthermore, the Human in the Loop (HITL) role is critical to refine AI generated ontology structures, take care of ethical considerations and enhance metadata alignment. Democratic expert validation through HITL adds yet more accuracy to knowledge graph and brings additional scalability challenges because of its reliance on human intervention.

Despite this, LLMs hold great promise for speeding up ontology development and enhancing knowledge representation, provided that they are fed with structured domain knowledge, have continuous human supervision and require fine tuning of prompting.

4.7 Conclusion

In conclusion, Ontology Construction using the larger Language Model (LLMs) has made it possible to automate the generation of taxonomy, as well as the extraction of relations and knowledge structuring. Both are able to process large numbers of datasets and discover meaningful relations that help accelerate the development of such knowledge systems across an enormous range of domains, including finance, healthcare and e-commerce.

Research reveals that organizing data with ontologies makes it easier to find and combine information. By adding business information to a structured framework, the ontology improves understanding of data meaning and generates better decision outcomes from available data. Validation and data preparation before building an ontology become crucial steps during evaluation and ontology development. Even though developing an ontology takes effort due to hard relationship modeling and complete data representation requirements the system's data connection advantages and better decision support make up for these problems. Future work will include improving the ontology using machine learning tools while expanding its contents and making it work better with large-scale systems.

Although LLMs do not suffer from the limitations of logical reasoning, hierarchical organization, and semantic coherence, it is often the case that the LLMs produce inconsistencies that need expert validation. Both structured prompting and Human in the Loop (HITL) methodologies help minimize these problems in a more accurate and ethical way. At some point, this automation becomes less efficient; however, the efficiency is not the main issue and the scalability comes down to balancing automation and the amount of human involved when needed. Furthermore, fine tuning LLMs in domain specific setting and LLMs in training using reinforcement learning provide areas of potential for improvement to the performance of LLMs in ontology construction.

Leveraging LLMs also saves a lot of time and effort in the ontology creation, however, LLMs are not fully replacing human domain experts. In the future, AI automation with a human validation hybrid model will be necessary to refine the knowledge representation and trust in the ontologies generated by AI. Future research directions are to improve the logical inference capability, reduce bias, and develop more structural prompting techniques to further strengthen the knowledge frameworks in LLM, so we can achieve more reliable ontology construction processes.

CHAPTER V:
DISCUSSION

5.1 Discussion of Results

Large Language Models (LLMs) are considered to be powerful ontology building infrastructure/resource for fast, on-demand, easy and scalable development in domain specific knowledge (Giglou et al., 2023). Because they can process tremendous amounts of unstructured data and discover the relationships from that data, they have a lot of value in domains such as healthcare, finance, and engineering as a whole. In comparison to the highly manual and expensive ontologies building approaches, LLMs are a more convenient, time, and resource efficient means for generating structured representations of knowledge from textual sources (Mahadevkar et al., 2024). However, even though they have potential, there are some issues to be resolved to make ontology construction reliable and effective.

The main worry regarding LLMs in ontology development is that, by nature, LLMs are biased and contextually inaccurate. In training LLMs on large scale datasets from the internet, LLM may inherit biases in the data and thus have ontological representations that either represent biases or are misleading (Neuhaus, 2023). These may bias the classification of concepts, relationships and hierarchical structures in an ontology. Since the reference basis matters in certain domains, the understanding of domain specific nuances is still hard for the LLMs. Constructed ontologies without any oversight of their outputs may be inaccurate with inconsistent results (Fahad et al., 2008).

In order to address these challenges, the role of Human in the Loop (HITL) approach in refining LLM generated ontologies is invaluable. This allows organizations to integrate expert intervention at different stages, for example, data preprocessing and constructing the knowledge structures to obtain higher accuracy and relevance of the constructed knowledge structures (Casado-Mansilla, 2024). Erroneous classifications, ambiguous relationships, and concepts generated can be cleaned up by human experts, ambiguous relationships clarified, and existing domain knowledge concepts confirmed. Therefore, it is not only more reliable to LLM when constructing the ontology, but also performs the process iteratively to ensure the depth and context of the information (Ling et al., 2023).

Additionally, prompt engineering, knowledge graph integration and iterative refinement can further improve the application of LLM to ontology development. Prompt engineering has the potential to allow researchers to design very specific, very specific queries that will encourage LLMs to produce more ontological structures, more relevant (Makin, 2024b). At the same time, LLMs can work better with knowledge graphs for additional consistency, as structured information can be used to verify and affirm produced relationships. The LLM achieves higher ontology quality through iterative refinement which involves receiving expert feedback during its iterative self-improvement process (Patil & Gudivada, 2024).

In essence, it seems that the role of LLMs in ontology construction would benefit from improvement with the advancement of LLMs, particularly in contextual understanding and adaptive learning. Future models should include phenomena that are self-corrective, bring context and have enhanced reasoning capabilities to produce better and more reliable ontology. Ethical considerations such as the requirement of (transparency, fairness, accountability) of the ontology building process will also be a determinant to responsible use of LLMs for knowledge representation (He et al., 2023).

From a business perspective, integrating predictive ontologies into operational processes can have transformative effects. Structured and accurate domain knowledge allows organizations to have better means to generate precise forecasts and data driven insights. It has been found that companies that use the most advanced predictive analytics can also cut through the decision making process and operational cost time by a significant amount, which can improve the overall efficiency of a company (Benbya et al., 2020; Brynjolfsson & McAfee, 2017). In terms of the accuracy of their predictions, the improved accuracy directly contributes to streamlined business operations in terms of faster, more informed decision making and reduced need for extensive manual intervention towards business operations.

Additionally, accurate, ontology-based predictions also provide the benefit of the ability to reduce cycle times in business process. In organizations which heavily rely on data driven strategy, integrating the LLM enhanced ontology can reduce lead length time to incorporate knowledge and validation resulting on expedited reaction to market shift. Such improvements are studied in terms of operational efficiency, and it is shown that they improve resource allocation, and increase competitiveness (Khare et al., 2022; McAfee & Brynjolfsson, 2012). In particular, these benefits are most apparent in areas like supply chain management and financial risk assessment where quick and correct information is crucial to stay at par with the competition.

Modern ontologies can finally be used strategically by businesses to improve efficiency and quality of production. Providing decision makers access to accurate and relevant data in a way that enhances their ability to understand the current and future state of the company's goods and services can help companies improve their goods and services. Based on the larger body of research on digital transformation, this strategy is supported by the fact that predictive modelling and integrated analytics are key for long term company success (H. Chen et al., 2012; Porter & Heppelmann, 2014). Put differently, LLM continuing to fuse LLM capabilities with human expertise is a strong framework, built for addressing these technological challenges while bringing greater real business benefits like that of reduced cycle times, increased efficiency and superior output quality.

5.2 Discussion of Research Question One

RQ1: What are the strengths and limitations of large language models (LLMs) in ontology construction?

LMs have transformed the process of ontology construction through LLMs for knowledge extraction, structuring, and organization, as well as automation of domain specific knowledge. On the other hand, what makes them strong are their ability to process an enormous amount of unstructured textual data and discover interrelations between concepts with little human intervention (Giglou et al., 2023). The automation decreases the time and resources required to develop ontology compared to manual methods. Moreover, LLMs are highly adaptable across various domains, making it possible to rapidly build knowledge structures for many deployments such as healthcare and finance as well as engineering (Uschold & Gruninger, 2022). They can understand natural language to interpret complex terminologies in the context, which improves the efficiency of the knowledge representation.

Nevertheless, LLMs also have their limitations as well. The inherent bias is one of their critical challenges because it is sourced from the large-scale data used to train. Thus, these biases may contribute to unreliable concepts classification and mapping that inaccurately form and distort the ontology's reliability (Bender et al., 2021). In addition, LLMs fail to achieve domain-specific precision and cannot generally achieve the contextual depth needed for specialized fields, in general. Probabilistic outputs make them rely on human expertise for validation. A further limitation is their inability to reason logically beyond their training data so that they can generate plausible but incorrect associations (Trujillo, 2023).

Integrating a HITL approach is necessary to mitigate these weaknesses to refine and validate LLM generated ontologies to gain higher accuracy and contextual relevance (Memarian & Doleck, 2024). In terms of effectiveness, precision and reliability of LLM based on ontology construction can be further improved by means of prompt engineering, knowledge graph integration, and iterative model refinement (Nickel et al., 2016). Future applications of LLMs with contextual adaptation and self-learning capabilities have the promise to overcome the existing challenges. The responsible use of LLMs in ontology development will also involve the consideration of ethical factors like transparency and fairness in model training (Floridi & Chiriatti, 2020).

From a business point of view, the operational improvements that can be realized through LLM enhanced ontology construction make it a promising approach. Ontologies in structures give a more accurate prediction by providing a clear and context rich view of the domain knowledge. It enables efficiency gains through improvement in turnaround times in decision making processes and waste of resources. The research on predictive analytics suggests that the organizations that correctly align with data driven insights have decreased cycle times, decreased operational costs, better quality products and services, high customer loyalty and reduced inventory costs (Alqudah & Muradkhanli, 2021; Brynjolfsson & McAfee, 2017). Since rapid response is critical in business environments in which supply chain management or financial risk assessment are involved, the ability to rapidly produce and validate reliable ontologies directly translates into improved competitiveness and strategic agility.

In addition, the integration of refined ontological structures into business processes leads to the generation of higher quality outputs. Strongly structured and accurate data are key to effective forecasting and risk management resulting in better quality product and service quality delivery. Research on operational efficiency has shown that it leads to faster decision making in addition to better decisions with lowered errors in and better data consistency in outcomes through the use of smarter predictive models (CHABANET et al., 2024; McAfee & Brynjolfsson, 2012). A strategic point of investment in the robust LLM supported knowledge management framework is reflected by the synergy between technology and business process optimization.

Finally, even though LLMs offer a huge promise for automating ontology construction, their full capacity comes with the help of human expertise and business-oriented enhancements. The evolution of these systems, guided by iterative refinement and ethical oversight, will pave the way for substantial improvements in operational efficiency and output quality, thereby offering tangible benefits to modern businesses.

5.3 Discussion of Research Question Two

RQ2: How can LLMs be effectively integrated into ontology construction to improve efficiency and scalability?

This integration of LLMs into ontology construction has made ontology building very highly efficient and scalable by allowing automation of the knowledge extraction, classification, as well as structuring processes. The process of building traditional ontology is labor intensive and requires a lot of domain expertise to construct entities, relationships and hierarchies. Due to being trained on large amounts of textual data, LLMs offer automatic generation and updates of ontologies by retrieving semantic relationships, reducing manual labor and increasing adaptability. New information can automatically update ontologies because they interpret unstructured data thereby enhancing their value in healthcare and finance applications alongside technology use.

The ability of LLMs for entity recognition and relation extraction is a primary advantage of using LLMs for ontology development as they are essential for building complete ontological structures. In contrast to the rule based and statistical approaches, LLMs use contextual embeddings to differentiate fine patterns of relationship and consequently lead to more flexible and scalable ontology construction (Olga Perera, 2024). In addition, their ability to cross-domain learning enables them to exchange knowledge between ontologies from different fields and to integrate knowledge (Fumagalli et al., 2020). The fusion of the transformer architectures (GPT, etc.) makes LLMs able to deal with a large corpus rapidly and reliable with high contextual accuracy for ontology development (Ibrahim et al., 2024).

Nevertheless, integrating LLMs into ontology construction has its own problems. Reliability for LLM generated ontologies is one critical issue because models can output inconsistent or biased outputs because of the data they are trained on (Saki Norouzi et al., 2024). In addition, LLMs experience domain-specific precision, and especially in the technical or specialized fields where expert validation is not an option (Padiu et al., 2024). On the other hand, LLMs possess another limitation: they are not explainable; in other words, they can infer relationships well, but their reasoning processes are not transparent to experts, and they cannot verify or refine them.

In order to fully leverage LLMs for ontology building, the hybrid models, that combine AI driven automation with HITL approaches (Zheng et al., 2017) have been on the way. Expert validation integration into the LLM workflow aids organizations to improve ontology outputs in terms of accuracy and credibility. Consequently, domain specific datasets can also be further fine-tuned with LLMs to achieve higher precision, and combining this with higher precision generated ontologies that are domain specific to a specialized field improves the precision of such generated ontologies (Delgaty et al., 2024).

Moreover, semantic reasoning techniques and integration with both the LLMs and structured knowledge bases can reduce the biases and make the ontology generation consistent. For adoption, Explainable AI (XAI) methods are needed, since they can support bridging the LLSM driven automation gap and the interpretability of humans at the domain level for them to effectively refine ontology structures based on expertise (Wolniak, 2023).

In addition to the technical merits, integrating LLMs in ontology construction provides great business benefits. Ontology generation aids enhanced predictive analytics, improving the operation efficiency and decreasing cycle times. However, in order to benefit from using data for strategic decisions, businesses lack the ability to generate structured knowledge at lightning speed. In predictive analytics research, data driven firms can substantially achieve efficiency gains by reducing manual processing and operations cost (H. Chen et al., 2012). For instance, predictions should be precise and timely in the areas like supply chain management or financial risk management. Ontology updating through automation enables companies to quickly and rapidly respond to the changes in market, thereby reducing turnaround times and increasing competitiveness.

In addition, the refined ontological structures are directly integrated into improved quality of outputs. This is more than just a matter of semantics. When data is high fidelity (precise and contextual), and decision makers have access to it, they have a clearer picture of the trends, they are better able to foretell the future and do what's needed, both to eke out an optimistic outcome and to mitigate the worst. It has been demonstrated that, by adopting advanced predictive models, not only are decisions made more quickly, but they are made better with superior outcomes since errors are minimized and processes become more dependable (Jofre, 2011). It aids in a higher quality output that helps to ensure strategic growth and improve customer satisfaction. The ability to perform with such capabilities, however, is a critical edge to help guarantee that organizational decisions are based on robust, data driven insights in highly competitive markets.

The strategic use of LLM to modify ontology construction is also a perfect fit to other digital transformation efforts. In today's economy it's not uncommon to see businesses that will use the scalability of automated systems, to provide flexible, and lower operational overhead while integrating it into their operations. Combining LLM automation and human oversight means the ability to continuously update and improve ontologies without sacrificing accuracy. The continuous improvement cycle results in more reliable predictions, more efficient, and thus a better quality of service and output in the various areas of business. Ever increasing the power of the integrated AI systems, as organizations harness this power, there is a sustained competitive advantage and operational excellence.

5.4 Discussion of Research Question Three

RQ3:* What *are the key steps in pre-processing OSHA accident and injury data, and how does this impact the consistency and quality of input for ontology construction?

Before further processing, data preprocessing is an important step for deal with OSHA accident and injury data with consistency, completeness and accuracy. OSHA datasets contain raw data from different sources, which creates problems including the missing values, inconsistent formatting, duplicates and erroneous records (J. Y. Lee et al., 2020). Proper preprocessing of data and reliability of the insights derived from data mitigate these challenges.

The results on the analysis of the OSHA Accident and Injury dataset provide the practical and useful information on workplace hazards, data structuring with ontologies and quality of relationship extraction procedures. It discusses key such as findings from exploratory data analysis, ontology generation, evaluation metrics, and refinements of this work. The EDA phase identified dominant workplace hazards and critical incidents using visual methods such as word clouds. The ontology generation process demonstrated a structured approach to extracting relationships from incident descriptions. GPT-4 effectively categorized incidents into entity-relationship models, as depicted above. The structured relationships, such as "employee -> fell from -> ladder" and "fall -> resulted in -> death," confirmed that fall-related incidents were a major concern in workplace safety. Additionally, Figures provided a frequency-based analysis of the most common relationships in workplace injuries. The high occurrence of "fall -> resulted in -> killed" and "fall -> resulted in -> injury" relationships further validated the significance of fall prevention strategies. The ontology relationship in results illustrates the interconnections between different workplace hazards, highlighting the "employee" as the central entity in most incidents.

The automated ontology evaluation indicated robust performance across key metrics. The system achieved a 0.69 accuracy rate in correctly extracting relationships compared to the refined ontology. On average, 2.46 relationships were extracted per record, demonstrating moderate coverage of incident descriptions. A relevance score of 0.78 confirmed that most extracted relationships aligned with domain-specific safety rules. The ontology achieved 1.00 consistency by eliminating duplicate relationships, ensuring data integrity. However, several issues were identified and subsequently corrected to enhance the ontology's quality. To maintain uniformity, standardized in lowercase formatting was applied. Examples such as 'employee' were replaced with terms like 'injured employee.' Last, ambiguous terms such as "experienced" were made more specific as "suffered", and "occurred on" was changed to "occurred from", to clear things up. Standardized relationship definitions were made for readability purposes and efficiency during computation. Entities were related to others, i.e. "firework" and "incident" were merged into "firework accident."

This study discusses the results and draws attention to the potential of ontology-based approaches for structuring work place safety data. Through the extracted relationships, explanatory insights of accident patterns can be used in better risk assessment and safety intervention. Yet, evaluation metrics indicate that more accuracy and completeness need to be increased. Future work also consists of integrating machine learning techniques for automating ontology refinement, and increasing the size of the dataset with more workplace incidents. Also, the ontology can be enhanced with interoperability with the safety management systems with the ability to maintain real-time hazard monitoring and predictive analytics on the workplace safety. Overall, it is shown that a methodology based on an ontology effectively organizes and extracts knowledge from workplace incident reports. This study increases the utility of workplace safety analytics, improving upon extracted relationships and data structuring in order to enhance the decision making and proactive hazard prevention strategy.

Accurate preprocessing and structured ontology construction not only enhance data quality but also have significant business implications. Organizations that adopt such methodologies can derive more accurate predictions from their data, which in turn lead to faster decision-making processes and reduced turnaround times. Businesses can use robots to extract data and maintain relationships in real time, such that they can quickly update knowledge base and respond to emergent workplace hazards with speed. This has shown out in studies that integration of automated analytics with advanced data preprocessing techniques have eliminated manual processes and cut down operational costs as well as reducing manual processing time (Davenport & Ronanki, 2018).

Processing the data in a meticulous manner directly contributes to improving output quality in predictive models through improved data quality. A more reliable ontology can help the predictive analytics to provide better insights into the patterns of accident and risk factors that will improve the proactive hazard prevention. One example is that accurately extracted relationships of workplace incidents enable organizations to identify future risks and reallocate resources to increase safety standards and decrease the incidence rate. Reactive and proactive decision making is crucial in such high stakes environments such as Industrial Safety Management and therefore, high quality structured data that supports that is critical.

Ontology based approach to preprocess OSHA data from a strategic perspective provides competitive advantage. These refined ontologies can be integrated with safety management systems of companies for real-time hazard monitoring and predictive analytics. This integration enhances the accuracy of risk assessments and supports continuous improvement in workplace safety protocols. Not only can incident prediction and prevention reduce operational disruptions, but it also helps bring down insurance rates and boosts business reputation. One of the most important factors in a company's long-term success is its capacity to draw useful conclusions from data in a timely manner (Porter & Heppelmann, 2014).

5.5 Discussion of Research Question Four

RQ4: How has the human-in-the-loop (HITL) approach been applied in ontology and knowledge graph construction?

To improve the accuracy, interpretability, and adaptability of ontology and knowledge graph (KG) construction, human expertise has been integrated with automated techniques through the HITL approach. Despite the advances of ML and NLP models in ontology generation on a large scale, human involvement continues to be a necessity to address the ambiguities, validate relationships, and guarantee the domain specific precision (B. Zhang et al., 2023a). When entirely automated ontology creation does not work well with possible errors in entity recognition, classification, and semantic relationships making up the knowledge representation, HITL offers an alternative solution (Zengeya & Vincent Fonou-Dombeu, 2024).

Knowledge validation is one of the main use cases of HITL, as it allows experts to review and modify automatically extracted relationships and classifications. For example, human intervention can validate disease-symptom associations extracted from text mining and ML based extractions in biomedical ontologies to conform to established medical taxonomies (Simmons et al., 2017). Finally, during the iterative refinement process, the ML models collaborate with the domain experts to add missing concepts, merge duplicate entities and correct erroneous classification (Xie et al., 2024).

HITL is another important application within its analysis of the problem of ontology alignment, where there are multiple heterogeneous knowledge sources to be integrated. Terminology inconsistencies are notoriously difficult to cope with for automated techniques, and human experts are needed to reconcile the differences in datasets (Tocchetti & Brambilla, 2022). This is accomplished by the human review of contextual understanding in industry specific ontologies such as legal and financial knowledge graphs, where the advantage of human reviewers over automated systems in terms of quality of entity linking and relationship extraction is proven (Peng et al., 2023).

HITL is applied in semi-supervised ontology learning to label a small part of training data, which instructs ML models to learn better representations. Annotating and validating ontology elements is done using crowdsourcing platforms, i.e., Amazon Mechanical Turk, improving scalability while maintaining accuracy (Washington, 2022). This approach has proved useful in order to build large-scale domain ontologies, for example in e-commerce and social media knowledge graphs (Dessì et al., 2021). In addition, HITL is essential to bias mitigation and ethical oversight during knowledge graph construction. Despite the best efforts of developers, biases that defined the training data can be unwittingly propagated to skew or distort ontological representations. Such iterative feedback loops can incorporate human oversight and help identify and correct for more ethical and balanced ontology structure (Madubuko et al., 2024).

Overall, the use of HITL greatly improves ontology and KG construction, as the data quality is improved, errors are reduced, and contextual relevance is ensured. However, the legal, ethical, and human expertise dimension can never be automated away, despite the fact that automation speeds up the extraction of large-scale knowledge (Mpofu, 2023). Future research should be aimed at improving efficiency and human effort in HITL frameworks, based on deploying interactive AI systems that are dynamically adaptive to the expert feedback (Casado-Mansilla, 2024).

Beyond technical improvements, integrating HITL within ontology construction has profound business implications. In the realm of operational efficiency, the ability to integrate human expertise into automated processes significantly reduces cycle time and improves decision accuracy. Accurate and validated ontologies empower organizations to make predictions that are both reliable and timely. This improvement in prediction accuracy translates into faster decision-making processes, as businesses can rely on well-structured and refined data for risk assessment and strategic planning. As an example, in the industries like finance and healthcare, faster turnaround time in data processing is directly associated with faster response to emerging risks and opportunities.

In addition, the output quality is improved with HITL enhanced ontologies. /error in entity extraction and relationship mapping are minimized by using expert validation and as such more precise and actionable insight is produced. Accurate predictive models are fundamental to driving business efficiency by being proactive at interventions, and this precision is necessary for developing them. Predicting incidents of these events accurately in operational environments costs less operationally and quality of the service is improved. The accuracy of predictions also means that the quality of output can be enhanced, businesses can optimize resource allocation, cut downtime and thus gain an edge in the market. The paired t-test analysis confirmed statistically significant improvements in accuracy, consistency, and relevance following the HITL refinement process ($p < 0.05$). Completeness did not exhibit a statistically significant change, indicating that the LLM output was already reasonably complete, and HITL primarily enhanced the correctness and semantic quality. The expert questionnaire results further support these findings, with average ratings above 4 for all metrics, highlighting substantial perceived improvement post-HITL.

In the end, HITL is strategically integrated with an automated ontology generation to form a robust feedback loop which continues to refine data quality. The surface of this is that human expertise and ML automation will dance back and forth, causing knowledge systems to stay able to adapt to changing conditions, thus enhancing the scalability and effectiveness of decision-making processes. These enhanced systems offer organizations the opportunity to reduce manual intervention and ‘error’, and increase the reliability of their predictive analytics. This means that businesses are not just prepared for the current challenges, but are also in a stronger position to prepare the business for the future growth and innovation.

5.6 Discussion of Research Question Five

RQ5: What methods can be used to identify and resolve inconsistencies, redundancies, and ambiguities in a generated ontology, and how can human intervention improve its accuracy?

LLMs and Machine Learning (ML) help the construction process for ontologies as human oversight remains important to create consistency, eliminate repetitions and resolve uncertainties while the automation progresses. In the end, the ultimate result of the automatic ontology generation process is semantic inconsistencies and unclear conceptual relationships caused by the insufficient understanding of the context and the domain specific requirements (Ozkaya, 2023). These problems can be resolved by requiring human reviewers to conduct checks, which results in better ontology accuracy and usability within knowledge graph applications and makes possible semantic search and AI-driven decision-making processes.

- **Identifying and Resolving Inconsistencies**

The main issue with automatically generated ontologies involves logical inconsistency between conflicting or contradictory relationships. Differences in terminology and incorrect entity classifications, together with ML model interpretation errors cause inconsistencies to appear within constructed ontology networks (Ma et al., 2021). Human experts must step in for ontological rule assessment together with the identification of any existing issues. Ontology debugging techniques, together with reasoning-based validation procedures, permit experts to find and fix contradictions that preserve ontology structural coherence.

- **Eliminating Redundancies**

During automated ontology production multiple concepts along with relationships tend to duplicate since datasets overlap while terms may differ from one another. An ontology may present duplicate versions of the same entity ("AI Ethics" and "Ethical AI") which results in search and logical processing issues. The process of ontology merging and alignment requires human expertise to execute semantic similarity measures (Heyder et al., 2023) which helps combine duplicate concepts while optimizing the ontology framework. Crowdsourcing methods together with evaluations from domain specialists work to solve these duplicate entries (Thuan et al., 2018).

- **Resolving Ambiguities**

The ontology development by LLMs can emerge semantic ambiguities that produce multiple understandings of entities or relationships. The main causes of ambiguities arise from two conditions: first, from polysemous terms, and second, from complex relationships between concepts that are unclear. When using the word "bank" people must determine whether it means a financial institution or the edge of a river because the definition remains unclear without context. The necessary domain expertise from human experts enables them to improve ontology definitions and metadata tags while applying word sense disambiguation methods (Schadd & Roos, 2015).

- **Improving Ontology Accuracy Through HITL Approaches**

Software development methods involving HITL enable experts to validate ontologies during multiple stages of ontology creation. The HITL framework allows experts to make systematic development improvements through repeated human review and annotation of machine processing outcomes (Takerngsaksiri et al., 2024). Expert curators from biomedical research help validate disease-drug interaction suggestions through NLP models, according to Li et al., (2021). Documentation and human inspections that use machine learning for verification help refine ontologies successfully, according to Tiddi & Schlobach, (2022).

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary of Findings

This research developed an ontology method to make workplace safety assessments better using OSHA Accident and Injury data. The research project used GPT-4 and other

Large Language Models (LLMs) to create an ontology system for workplace safety. Our method analyzed workplace incident texts to develop a safer work environment evaluation system. Our process started by preparing and studying workplace incident text records to find important elements of safety hazards and connections related to workplace harm. GPT-4 transformed the project elements into an organized structure through its ontology generation process. With automated analysis, this system found regular workplace perils and failure sequences plus safety improvement answers to drive workplace protection assessments better.

The study tested whether the created ontology could improve safety reviews and found valuable differences between it and traditional evaluation methods. The researchers used metrics to measure how well the ontology met standards for being easy to understand and also containing all needed information. The research proves that LLM-based ontology creation helps safety expert's complete workplace safety checks in less time while enhancing their ability to identify hazards before they cause issues.

The research showed that using an ontology system works well when LLMs extract workplace safety data from free-text incident stories. Our research results show the following key findings:

1. **Ontology Generation:** A structured methodology was developed for ontology extraction from workplace incident descriptions, incorporating several NLP techniques such as text normalization, tokenization, and lemmatization. These pre-processing steps were critical in standardizing textual data, ensuring that variations in terminology did not impact the consistency of ontology generation. GPT-4 was instrumental in extracting vital entities together with their relationships and hazard categories from incident report texts.

2. **Data Pre-processing:** Data preparation enriched the quality of our discovered information through effective clean-up methods. Our approach of removing stop words and fixing text when needed both made select statements directly linked to work hazards less repetitive and more accurate. The study established a reliable workplace danger database using text cleaning to make sure its output matches official records.
3. **Ontology Evaluation:** Our generated ontology needed an evaluation based on four metrics including accuracy, completeness, relevancy and consistency. Our results show that automated extraction works well with actual workplace safety records since it produces correct results 69% of the time. The LLM-based construction system produced accurate workplace safety models that use relevant information to help organizations make better safety decisions. Research results outlined specific ways to enhance entity connection definitions and context interpretation within the system.
4. **Exploratory Data Analysis (EDA):** The analysis methods in Exploratory Data Analysis revealed more detailed information about work-related dangers and their corresponding accident patterns. The combination of word clouds and bar charts through visualization tools enabled stakeholders to see what work hazards appeared most frequently and the types of injuries along with dangerous areas in the workplace. The findings from this analysis helped enhance knowledge about safety hazards in workplaces which led to better data-based choices in handling occupational health issues.
5. **Human Review and Refinement:** The ontology construction process experienced speed-up benefits from automation but manual assessment served as essential for obtaining greater precision. To ensure accuracy in capturing real-world workplace

safety concepts human experts performed reviews of entity naming as well as formatting definitions and relationship specifications. A deployment model combining artificial intelligence automation with professional validation created practical workplace safety applications through better performance of the final ontology.

6.2 Implications

Our findings show why ontology improvement is essential for ensuring accurate structured data representation and suitable use in different applications that need structured knowledge systems. Standardizing relationships, removing redundancies, and resolving ambiguities, the refined ontology helps for automatic reasoning, interoperations among systems, and increasing decision accuracy. The study shows that small word changes such as replacing 'experienced' with 'suffered' and some refined 'on' to 'from' makes the ontology more contemporary and worse. This reinforces the fact that reality and knowledge modelling necessitate careful word selection.

Human in the Loop (HITL) integration in this approach emphasizes the need for human expertise in validating and refining ontology structures in the areas where such ethical considerations, besides the need for regulatory compliance, also impose the need for domain-specific accuracy. Safeguarding against errors like misinformation, security risks, or biased decisions based on machine-generated ontologies, HITL completes the promise of machine-generating ontologies. While the study also shows these key challenges—scalability of HITL, reliance on expert validation, and time-consuming manual interventions—overall, it can be reused and adapted for further research towards HITL, leveraging all its benefits and mitigating the drawbacks. These challenges serve as a call for the adoption of a hybrid solution, which is a combination of automation and human supervision using machine learning methods and crowd validation to ease the process of ontology refinement without sacrificing accuracy and reliability.

Additionally, it is shown through the results of experimental evaluation that automated ontology models have a 69% accuracy rate for relationship extraction and improve completeness and consistency. Through the refinement process, the precision of relationships was successfully increased, and redundancy was removed, resulting in a better coherent and structured knowledge system. This finding indicates that the systematic standardization of ontology by human action can increase the usability of this artifact when used for knowledge-based applications, for instance, natural language processing, decision-support systems and data integration platforms.

This study demonstrates that the HITL approach effectively enhances ontology quality. Both quantitative (t-test) and qualitative (expert survey) evaluations confirm improvements in accuracy, consistency, and relevance, while completeness remained stable. The combined analysis validates HITL as a reliable method for improving AI-generated ontologies. The inclusion of a diverse expert panel further strengthens the reliability and generalizability of these findings.

6.3 Recommendations for Future Research

Several areas for further research based on the findings and implications of this study should be pursued to improve the refinement of ontology and knowledge graph construction.

1. **Future research of Scalability of Human in the Loop (HITL) Approaches** – The HITL approach to End-User Development (EUD) should focus on frameworks that offer a balance between the level of automation and the level of human oversight. Analysing ontology accuracy depends on active learning

combined with reinforcement learning as well as crowdsourced validation to decrease the necessity of expert-driven interventions.

2. **Improvement in further automata Ontology –** An improved understanding of machine learning and Natural Language Processing (NLP) can make further automata Ontology Automation. In fact, AI should be used to develop mechanisms for real-time detection of ambiguities, inconsistencies and redundancies, thus reducing the need for manual validation.
3. **Develop Methods of Ensuring Cross-Domain Ontology -** Future studies should develop methods of ensuring cross-domain ontology interoperability. It should develop universal standards and frameworks that will allow integration across industries without problems such as healthcare, finance, cybersecurity and environmental sciences.
4. **Improving Precise and Recall of Relationship Extraction Models -** The precision and recall of the relationship extraction models should be improved through the addition of more advanced linguistic and contextual analysis. The disambiguation of entities and refinement of taxonomic structures can increase the accuracy and usefulness of ontology relationships.
5. **Keeping Ontology Evaluation Metrics and Benchmarking –** The standardization and automation of ontology refinement can help to improve consistency and reliability at automation.
6. **Real-Time Data Processing with Ontology -** Future work could aim at integrating real-time data processing with ontology refinement to provide dynamic updates of ontology and thereby enhance adaptation to fast-changing environments like those of financial markets, cyber security threat detection, and crisis management.

6.4 Conclusion

Ontology refinement is crucial to the quality of representation in a knowledge graph, even more so in ensuring the accuracy and consistency of a knowledge graph. The study verifies the potential to optimize the ontology system clarity, precision and interoperability by addressing inconsistency, redundancies and ambiguities in a structured refinement process. A human-in-the-loop (HITL) approach provides an excellent method to combine automatic knowledge extraction with expert validation in order to increase the accuracy and contextual relevance of an ontology.

These refinements, for example, standardize terminology, eliminate duplicate relationships and restructure ambiguous definitions, facilitate the overall quality of ontology systems. It also emphasizes the value of the use of advanced NLP techniques, machine learning models and domain-specific rules to improve the relationship extraction and entity classification. Experimental results show that compared to the use of automated systems alone, used refinement of ontology models leads to a significantly increased level of precision, completeness and consistency.

At the same time, scalability, automation, and ethical issues of building ontologies are still issues. In future research of ontology automation, the cross-domain standardization and the real-time data integral, these challenges will be addressed to further advance the field. Through the refinement of ontology structures, they eventually become more applicable to various industries for more effective knowledge management, better decision making and better data-driven insights.

Conclusively, this research investigated how Large Language Models (LLMs), specifically GPT-4, operate in ontology-based workplace safety assessment. The research proved how industrial automation transforms raw workplace incident statements into defined ontological frameworks that improve both safety data extraction and organizational capabilities and retrieval processes. Text normalization combined with tokenization and lemmatization implemented through NLP techniques allowed the investigation to create an ontology that collects essential workplace hazard and risk factor and incident connection data. The methodology revealed that ontology creation with LLM support decreased human workload and improved operational speed and scored 69% precise in testing. Human experts must play a crucial role during the extraction process since it requires refinement and validation of generated knowledge. The analysis through Exploratory Data Analysis (EDA) confirmed structured analysis approaches for occupational health practice by revealing patterns between workplace hazards and incidents.

The practical along theoretical applications derived from this work provide meaningful benefits. The developed ontology functions as a component whose integration with workplace safety management systems allows organizations to make better risk evaluations, establish improved safety protocols, and improve their decision-making capability. This research enhances the field of AI-driven knowledge engineering by presenting a systematic evaluation methodology for ontology assessment that teaches the importance of keeping automation in harmony with expert human intervention. Future work should concentrate on improving ontology models with machine learning systems and broadening data sources for enhanced generalization capabilities as well as present-time ontology update mechanisms for maintaining contextual accuracy. Workplace safety evaluations require ethical developments that will address both bias reduction in LLM-generated ontologies together with clarity about their processes. Strategic hazard detection and workplace safety, along with better risk prevention, result from the use of expert-validated AI advances for structured knowledge extraction processes.

REFERENCES

Abolhasani, M., & Pan, R. (2024). *Leveraging LLM for Automated Ontology Extraction and Knowledge Graph Generation*. https://doi.org/10.48550/arXiv.2412.00608

Agarwal, V. (2015). Research on Data Preprocessing and Categorization Technique for Smartphone Review Analysis. *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2015907309

Aggarwal, T., Salatino, A., Osborne, F., & Motta, E. (2024). *Large Language Models for Scholarly Ontology Generation: An Extensive Analysis in the Engineering Field*.

Agrawal, G., Kumarage, T., Alghamdi, Z., & Liu, H. (2023). Can knowledge graphs reduce hallucinations in llms?: A survey. *ArXiv Preprint ArXiv:2311.07914*.

Agüera Y Arcas, B. (2022). Do Large Language Models Understand Us? *Daedalus*. https://doi.org/10.1162/DAED_a_01909

Al-Turki, D., Hettiarachchi, H., Gaber, M. M., Abdelsamea, M. M., Basurra, S., Iranmanesh, S., Saadany, H., & Vakaj, E. (2024). Human-in-the-Loop Learning with LLMs for Efficient RASE Tagging in Building Compliance Regulations. *IEEE Access*, *PP*, 1–1. https://doi.org/10.1109/ACCESS.2024.3512434

Al-Turki, D., Hettiarachchi, H., Medhat Gaber, M., Abdelsamea, M. M., Basurra, S., Iranmanesh, S., Saadany, H., & Vakaj, E. (2024). Human-in-the-Loop Learning With LLMs for Efficient RASE Tagging in Building Compliance Regulations. *IEEE Access*, *12*, 185291–185306. https://doi.org/10.1109/ACCESS.2024.3512434

Aliero, A., Bashir, S., Aliyu, H., Tafida, A., Kangiwa, B., & Dankolo, N. (2023). Systematic Review on Text Normalization Techniques and its Approach to Non-Standard Words. *International Journal of Computer Applications*, *185*, 975–8887.

Alijoyo, F. A., Janani, S., Santosh, K., Shweihat, S. N., Alshammry, N., Ramesh, J. V. N., & Baker El-Ebiary, Y. A. (2024). Enhancing AI interpretation and decision-making: Integrating cognitive computational models with deep learning for advanced uncertain reasoning systems. *Alexandria Engineering Journal*, *99*, 17–30. https://doi.org/10.1016/j.aej.2024.04.073

Alqudah, M. A., & Muradkhanli, L. (2021). Artificial Intelligence Applications That Support: Business Organizations and E-Government in Administrative Decision. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3866216

Alsobhi, A., & Amare, N. (2022). Ontology-Based Relational Product Recommendation System. *Computational and Mathematical Methods in Medicine*. https://doi.org/10.1155/2022/1591044

Amalki, A., & Bouzit, A. (2025). Deep Learning-Driven Ontology Learning : A Systematic Mapping Study. *Engineering, Technology & Applied Science Research*, *15*(1), 20085–20094.

Amershi, S., Cakmak, M., Knox, W. B., Kulesza, T., & Lau, T. (2013). IUI workshop on interactive machine learning. *International Conference on Intelligent User Interfaces, Proceedings IUI*. https://doi.org/10.1145/2451176.2451230

Amini, R., Norouzi, S. S., Hitzler, P., & Amini, R. (2024). Towards Complex Ontology Alignment using Large Language Models. *ArXiv*, *abs/2404.1*.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W., & Abbasi, H. M. (2018). A survey of ontology learning techniques and applications. In *Database*. https://doi.org/10.1093/database/bay101

Asmi Agarwal, A. A. (2024). Is AI the end of Human Creativity. *Journal of Advances in Science and Technology*, *20*(2), 21–27. https://doi.org/10.29070/9qgeet16

Babaei Giglou, H., D'Souza, J., & Auer, S. (2023). LLMs4OL: Large Language Models for Ontology Learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-031-47240-4_22

Baldazzi, T., Bellomarini, L., Ceri, S., Colombo, A., Gentili, A., Sallinger, E., & Atzeni, P. (2024). Explaining Enterprise Knowledge Graphs with Large Language Models and Ontological Reasoning. In A. Amarilli & A. Deutsch (Eds.), *The Provenance of Elegance in Computation - Essays Dedicated to Val Tannen* (Vol. 119, pp. 1:1--1:20). Schloss Dagstuhl -- Leibniz-Zentrum für Informatik. https://doi.org/10.4230/OASIcs.Tannen.1

Bamansoor, S., El-Ebiary, Y. A. B., Yusoff, F. H., Pathmanathan, P. R., Saany, S. I. A., Bakar, Z. B. A., Aseh, K., Pande, B., & Al Moaiad, Y. (2021). Evaluation of Chinese electronic enterprise from business and customers perspectives. *2021 2nd International Conference on Smart Computing and Electronic Enterprise: Ubiquitous, Adaptive, and Sustainable Computing Solutions for New Normal, ICSCEE 2021*. https://doi.org/10.1109/ICSCEE50312.2021.9498093

Barak, S., & Mokfi, T. (2019). Evaluation and selection of clustering methods using a hybrid group MCDM. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2019.07.034

Baumann, M., Weil, M., Peters, J. F., Chibeles-Martins, N., & Moniz, A. B. (2019). A review of multi-criteria decision making approaches for evaluating energy storage systems for grid applications. In *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2019.02.016

Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021). Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions. In *IEEE Access*. https://doi.org/10.1109/ACCESS.2021.3072900

Bavolar, J., & Bacikova-Sleskova, M. (2020). Decision-making styles and mental health—A person-oriented approach through clustering. *Journal of Behavioral Decision Making*. https://doi.org/10.1002/bdm.2183

Beg, A. (2024). *Use of NLP, AI and Ontologies for Software Specifications*. https://doi.org/10.13140/RG.2.2.10248.43527

Ben Abacha, A., Dos Reis, J. C., Mrabet, Y., Pruski, C., & Da Silveira, M. (2016). Towards natural language question generation for the validation of ontologies and mappings. *Journal of Biomedical Semantics*, *7*(1), 1–15. https://doi.org/10.1186/s13326-016-0089-6

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial Intelligence in Organizations: Current State and Future Opportunities. *Journal Article*.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3442188.3445922

Bernasconi, E., Ceriani, M., & Ferilli, S. (2024). LPG Semantic Ontologies: A Tool for Interoperable Schema Creation and Management. *Information*, *15*(9). https://doi.org/10.3390/info15090565

Biemann, C. (2005). Ontology Learning from Text: A Survey of Methods. *Journal for*

Language Technology and Computational Linguistics. https://doi.org/10.21248/jlcl.20.2005.76

Biloshchytskyi, A., Kuchansky, A., Andrashko, Y., Biloshchytska, S., Kuzka, O., Shabala, Y., & Lyashchenko, T. (2017). A method for the identification of scientists' research areas based on a cluster analysis of scientific publications. *Eastern-European Journal of Enterprise Technologies*. https://doi.org/10.15587/1729-4061.2017.112323

Boccacci, P., Beltramo, C., Sandoval Prando, M. A., Lembo, A., Sartor, C., Mehlenbacher, S. A., Botta, R., & Torello Marinoni, D. (2015). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Molecular Breeding : New Strategies in Plant Improvement*.

Boesch, G. (2024). *What Does it Mean to Have "Humans-in-the-Loop?"* Viso.Ai.

Booshehri, M., Emele, L., Flügel, S., Förster, H., Frey, J., Frey, U., Glauer, M., Hastings, J., Hofmann, C., Hoyer-Klick, C., Hülk, L., Kleinau, A., Knosala, K., Kotzur, L., Kuckertz, P., Mossakowski, T., Muschner, C., Neuhaus, F., Pehl, M., … Stappel, M. (2021). Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis. *Energy and AI*, *5*, 100074. https://doi.org/10.1016/j.egyai.2021.100074

Bornico, L., & Walden, I. (2011). Ensuring competition in the clouds: The role of competition law? *ERA Forum*. https://doi.org/10.1007/s12027-011-0223-y

Borrmann, A., König, M., Koch, C., & Beetz, J. (2018). Building information modeling: Why? What? How? In *Building Information Modeling: Technology Foundations and Industry Practice*. https://doi.org/10.1007/978-3-319-92862-3_1

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei,

D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*.

Brutzkus, A., & Globerson, A. (2019). Why do larger models generalize better? A theoretical perspective via the XOR problem. *36th International Conference on Machine Learning, ICML 2019*.

Brynjolfsson, E., & McAfee, A. (2017). *Machine, platform, crowd: Harnessing our digital future* (Vol. 4, Issue 1). WW Norton \& Company New York.

Building, A., Modeling, I., & Check, C. (2024). *Automated Building Information Modeling Compliance Check and Ontology*.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*. https://doi.org/10.1126/science.aal4230

Capellini, R., Atienza, F., & Sconfield, M. (2024). *Knowledge Accuracy and Reducing Hallucinations in LLMs via Dynamic Domain Knowledge Injection*. https://doi.org/10.21203/rs.3.rs-4540506/v1

Carnevale, A., Lombardi, A., & Lisi, F. A. (2024). A human-centred approach to symbiotic AI: Questioning the ethical and conceptual foundation. *Intelligenza Artificiale*, *18*(1), 9–20. https://doi.org/10.3233/IA-240034

Caruana, R., Elhawary, M., Nguyen, N., & Smith, C. (2006). Meta clustering. *Proceedings - IEEE International Conference on Data Mining, ICDM*. https://doi.org/10.1109/ICDM.2006.103

Casado-Mansilla, D. (2024). Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches. *Internet of Things (Netherlands)*, *25*(December 2023), 101048. https://doi.org/10.1016/j.iot.2023.101048

Caspari-Sadeghi, S. (2023). Artificial Intelligence in Technology-Enhanced Assessment:

A Survey of Machine Learning. *Journal of Educational Technology Systems*. https://doi.org/10.1177/00472395221138791

Çevikol, S., & Aydemir, F. B. (2019). Detecting inconsistencies of natural language requirements in satellite ground segment domain. *CEUR Workshop Proceedings*.

CHABANET, D., DO VALE, G., & WEPPE, X. (2024). Managing business model innovation in a changing environment: the case of SMEs in the pharmaceutical sector. *Management International*. https://doi.org/10.59876/a-3x4a-mzgx

Chang, Y., Wang, X., Wang, J., Wu, Y., Zhu, K., Chen, H., Yang, L., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P., Yang, Q., & Xie, X. (2023). *A Survey on Evaluation of Large Language Models*. https://doi.org/10.48550/arXiv.2307.03109

Charlwood, A., & Guenole, N. (2022). Can HR adapt to the paradoxes of artificial intelligence? In *Human Resource Management Journal*. https://doi.org/10.1111/1748-8583.12433

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly: Management Information Systems*. https://doi.org/10.2307/41703503

Chen, M., & Zhang, L. (2020). Human-in-the-loop Techniques in Machine Learning. *Machine Learning Journal*, 37–52.

Chen, N., Lin, X., Jiang, H., & An, Y. (2024). Automated Building Information Modeling Compliance Check through a Large Language Model Combined with Deep Learning and Ontology. *Buildings*, *14*(7). https://doi.org/10.3390/buildings14071983

Chen, S., Leng, Y., & Labi, S. (2020). A deep learning algorithm for simulating autonomous driving considering prior knowledge and temporal information. *Computer-Aided Civil and Infrastructure Engineering*.

https://doi.org/10.1111/mice.12495

Chen, X., & Watanabe, K. (2023). Ontology-driven AI: The hidden framework of enterprise success. *MIT Sloan Management Review*, *64*(3), 78. https://doi.org/https://doi.org/10.1016/j.mitsmr.2023.07.012

Chernyavskiy, A., Ilvovsky, D., & Nakov, P. (2021). Transformers: "The End of History" for Natural Language Processing? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-86523-8_41

Ciatto, G., Agiollo, A., Magnini, M., & Omicini, A. (2024). Large language models as oracles for instantiating ontologies with domain-specific knowledge. *ArXiv*, *abs/2404.0*.

Clemente, K. A. P., da Silva, S. V., Vieira, G. I., de Bortoli, M. C., Toma, T. S., Ramos, V. D., & de Brito, C. M. M. (2022). Barriers to the access of people with disabilities to health services: a scoping review. *Revista de Saude Publica*. https://doi.org/10.11606/s1518-8787.2022056003893

Collenette, J., Atkinson, K., & Bench-Capon, T. (2023). Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights. *Artificial Intelligence*. https://doi.org/10.1016/j.artint.2023.103861

Confalonieri, R., & Guizzardi, G. (2023). On the Multiple Roles of Ontologies in Explainable AI. *ArXiv Preprint ArXiv:2004.07213*.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*. https://doi.org/10.1023/A:1022627411411

Cortes, D., Leung, J., Ryl, A., & Lieu, J. (2019). Pharmacy informatics: Where medication use and technology meet. *Canadian Journal of Hospital Pharmacy*. https://doi.org/10.4212/cjhp.v72i4.2919

Davda, M. J., & V, S. D. (2024). Impact of Human-in-the-Loop Paradigm in AI Labeling and Annotation. *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, 1–4. https://doi.org/10.1109/CSITSS64042.2024.10817054

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*. https://doi.org/10.56726/irjmets42512

Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-Methods Research: A Discussion on its Types, Challenges, and Criticisms. *Journal of Practical Studies in Education*. https://doi.org/10.46809/jpse.v2i2.20

Deeba, K., Devi, O. R., Al Ansari, M. S., Reddy, B. P., Manohara, H. T., El-Ebiary, Y. A. B., & Rengarajan, M. (2023). Optimizing Crop Yield Prediction in Precision Agriculture with Hyperspectral Imaging-Unmixing and Deep Learning. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/IJACSA.2023.0141261

Delgaty, S., Lebang, E., & Lebang, E. (2024). *Boosting Domain Knowledge Understanding of LLMs through Fine-Tuning with a Novel KNN Algorithm*.

Dessì, D., Osborne, F., Reforgiato Recupero, D., Buscaldi, D., & Motta, E. (2021). Generating knowledge graphs by employing Natural Language Processing and Machine Learning techniques within the scholarly domain. *Future Generation Computer Systems*, *116*(September), 253–264. https://doi.org/10.1016/j.future.2020.10.026

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*.

Diligenti, M., Roychowdhury, S., & Gori, M. (2017). Integrating prior knowledge into deep learning. *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*. https://doi.org/10.1109/ICMLA.2017.00-37

Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. In *Computer Science Review*. https://doi.org/10.1016/j.cosrev.2021.100379

Dong, X. L. (2023). Generations of Knowledge Graphs: The Crazy Ideas and the Business Impact. *Proceedings of the VLDB Endowment*. https://doi.org/10.14778/3611540.3611636

Doukari, O., Wakefield, J., Martinez, P., & Kassem, M. (2024). An ontology-based tool for safety management in building renovation projects. *Journal of Building Engineering*, *84*, 108609. https://doi.org/10.1016/j.jobe.2024.108609

Doumanas, D., Bouchouras, G., Soularidis, A., Kotis, K., & Vouros, G. (2024). *From Human- to LLM-Centered Collaborative Ontology Engineering*. https://doi.org/10.1177/15705838241305067

Doumanas, D., Soularidis, A., Kotis, K., & Vouros, G. (2024). *Integrating LLMs in the Engineering of a SAR Ontology* (pp. 360–374). https://doi.org/10.1007/978-3-031-63223-5_27

Drori, I., & Teeni, D. (2024). Human-in-the-Loop AI Reviewing: Feasibility, Opportunities, and Risks. *Journal of the Association for Information Systems*, *25*, 98–109. https://doi.org/10.17705/1jais.00867

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. In *Human Resource Management Journal*. https://doi.org/10.1111/1748-8583.12258

Durmaz, A. R., Thomas, A., Mishra, L., Murthy, R. N., & Straub, T. (2024). An ontology-based text mining dataset for extraction of process-structure-property entities. *Scientific Data*, *11*(1), 1112. https://doi.org/10.1038/s41597-024-03926-5

Eastman, C., Lee, J. min, Jeong, Y. suk, & Lee, J. kook. (2009). Automatic rule-based checking of building designs. In *Automation in Construction*. https://doi.org/10.1016/j.autcon.2009.07.002

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., & Larson, J. (2024). *From Local to Global: A Graph RAG Approach to Query-Focused Summarization*. 1–15.

Edström, S. (2022). *Object detection for autonomous trash and litter collection*.

El-Sappagh, S., Franda, F., Ali, F., & Kwak, K. S. (2018). SNOMED CT standard ontology based on the ontology for general medical science. *BMC Medical Informatics and Decision Making*. https://doi.org/10.1186/s12911-018-0651-5

Estivill-Castro, V., Gilmore, E., & Hexel, R. (2022). Constructing Explainable Classifiers from the Start—Enabling Human-in-the Loop Machine Learning. *Information (Switzerland)*. https://doi.org/10.3390/info13100464

Fahad, M., Qadir, M. A., & Noshairwan, M. W. (2008). Ontological errors Inconsistency, Incompleteness and Redundancy. *ICEIS 2008 - Proceedings of the 10th International Conference on Enterprise Information Systems*, *2 ISAS*(January 2008), 253–258. https://doi.org/10.5220/0001696502530258

Fahad Mustafa, F. D. (2025). Deep Learning for Knowledge Representation : Automating Semantic Analysis Deep Learning for Knowledge Representation : Automating Semantic Analysis and Ontology Construction Date : January , 2025. *Resarch Gate*, *January*. https://doi.org/10.13140/RG.2.2.35145.45928

Farghaly, K., Amin, K., Mills, G., & Wilson, D. (2023). *Enhancing Interactions in Augmented Reality for Construction Sites: Introducing the Archi Ontology* (pp. 848–

855). https://doi.org/10.36253/979-12-215-0289-3.84

Farhana, A., Satheesh, N., Ramya, M., Ramesh, J. V. N., & El-Ebiary, Y. A. B. (2023). Efficient Deep Reinforcement Learning for Smart Buildings: Integrating Energy Storage Systems Through Advanced Energy Management Strategies. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/IJACSA.2023.0141257

Faria, C., Serra, I., & Girardi, R. (2014). A domain-independent process for automatic ontology population from text. *Science of Computer Programming*. https://doi.org/10.1016/j.scico.2013.12.005

Fathallah, N., Das, A., Giorgis, S. De, Poltronieri, A., Haase, P., & Kovriguina, L. (2024). NeOn-GPT : A Large Language Model-Powered Pipeline for Ontology Learning. *Research Gate*, *August*, 1–16.

Fernandez, I., Aceta, C., Fernandez, C., Torres, M. I., Etxalar, A., Mendez, A., Agirre, M., Torralbo, M., Pozo, A., Agirre, J. A., Artetxe, E., & Altuna, I. (2024). *Incremental Learning for Knowledge-Grounded Dialogue Systems in Industrial Scenarios*. 92–102.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. In *Minds and Machines*. https://doi.org/10.1007/s11023-020-09548-1

Friedman, R. (2023). Tokenization in the Theory of Knowledge. *Encyclopedia*, *3*, 380–386. https://doi.org/10.3390/encyclopedia3010024

Fumagalli, M., Conti, S., & Giunchiglia, F. (2020). *Ontology-Driven Cross-Domain Transfer Learning*. *October*.

Galkin, M., Yuan, X., Mostafa, H., Tang, J., & Zhu, Z. (2024). Towards Foundation Models for Knowledge Graph Reasoning. *12th International Conference on Learning Representations, ICLR 2024*.

GARBACZ, P. (2024). Large Language Models and Foundational Ontologies. *IEEE Security and Privacy*, *22*(2), 60–68. https://doi.org/10.1109/MSEC.2023.3345568

Gartner. (2024). *Top 5 Taxonomy Best Practices for Digital Transformation*.

Gascó, M. (2017). Living labs: Implementing open innovation in the public sector. *Government Information Quarterly*, *34*(1), 90–98. https://doi.org/10.1016/j.giq.2016.09.003

Gatt, A., & Krahmer, E. (2018). Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*. https://doi.org/10.1613/jair.5714

Giglou, H. B., D'Souza, J., & Auer, S. (2023). *LLMs4OL: Large Language Models for Ontology Learning*.

Gil, M., Albert, M., Fons, J., & Pelechano, V. (2020). *Engineering Human-in-the-Loop Interactions in Cyber-Physical Systems*.

Gómez-Carmona, O., Casado-Mansilla, D., López-de-Ipiña, D., & García-Zubia, J. (2024). Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches. *Internet of Things (Netherlands)*. https://doi.org/10.1016/j.iot.2023.101048

González-Eras, A., Santos, R. Dos, Aguilar, J., & Lopez, A. (2022). Ontological engineering for the definition of a COVID-19 pandemic ontology. *Informatics in Medicine Unlocked*. https://doi.org/10.1016/j.imu.2021.100816

Graber, M. L. (2013). The incidence of diagnostic error in medicine. In *BMJ Quality and Safety*. https://doi.org/10.1136/bmjqs-2012-001615

Grannis, S. J., Xu, H., Vest, J. R., Kasthurirathne, S., Bo, N., Moscovitch, B., Torkzadeh, R., & Rising, J. (2019). Evaluating the effect of data standardization and validation on patient matching accuracy. *Journal of the American Medical Informatics*

Association. https://doi.org/10.1093/jamia/ocy191

Guarino, N., Oberle, D., & Staab, S. (2009). What Is an Ontology? In *Handbook on Ontologies* (pp. 1–17). https://doi.org/10.1007/978-3-540-92673-3_0

Guizzardi, G., & Guarino, N. (2024). Explanation, semantics, and ontology. *Data & Knowledge Engineering*, *153*, 102325. https://doi.org/10.1016/j.datak.2024.102325

Habermann, M., Xu, W., Zollhofer, M., Pons-Moll, G., & Theobalt, C. (2020). DeepCap: Monocular Human Performance Capture Using Weak Supervision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR42600.2020.00510

Hadi, M. U., Al-Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M., Akhtar, N., Wu, J., & Mirjalili, S. (2023). *Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects*. https://doi.org/10.36227/techrxiv.23589741.v1

Haldorai, K., Kim, W. G., Pillai, S. G., Park, T. (Eliot), & Balasubramanian, K. (2019). Factors affecting hotel employees' attrition and turnover: Application of pull-push-mooring framework. *International Journal of Hospitality Management*. https://doi.org/10.1016/j.ijhm.2019.04.003

Hamie, J. M. K., & Abdul-Malak, M.-A. U. (2018). Model Language for Specifying the Construction Contract's Order-of-Precedence Clause. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *10*(3). https://doi.org/10.1061/(ASCE)LA.1943-4170.0000260

Haque, A. K. M. B., Arifuzzaman, B. M., Siddik, S. A. N., Kalam, A., Shahjahan, T. S., Saleena, T. S., Alam, M., Islam, M. R., Ahmmed, F., & Hossain, M. J. (2022). Semantic Web in Healthcare: A Systematic Literature Review of Application, Research Gap, and Future Research Avenues. *International Journal of Clinical*

Practice. https://doi.org/10.1155/2022/6807484

Hartmann, G., Shiller, Z., & Azaria, A. (2019). Deep reinforcement learning for time optimal velocity control using prior knowledge. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*. https://doi.org/10.1109/ICTAI.2019.00034

Harvard Business Review. (2022). *Taxonomies for Strategy: Structuring Organizational Knowledge*.

Hassan, F. ul, & Le, T. (2020). Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *12*(2). https://doi.org/10.1061/(ASCE)LA.1943-4170.0000379

Hassan, F. ul, & Le, T. (2021). Computer-assisted separation of design-build contract requirements to support subcontract drafting. *Automation in Construction*, *122*, 103479. https://doi.org/10.1016/j.autcon.2020.103479

Hassan, F. ul, Le, T., & Lv, X. (2021). Addressing Legal and Contractual Matters in Construction Using Natural Language Processing: A Critical Review. *Journal of Construction Engineering and Management*, *147*(9). https://doi.org/10.1061/(ASCE)CO.1943-7862.0002122

Hassanin, M., & Moustafa, N. (2024). *A Comprehensive Overview of Large Language Models (LLMs) for Cyber Defences: Opportunities and Directions*. 1–20.

He, K., Mao, R., Lin, Q., Ruan, Y., Lan, X., Feng, M., & Cambria, E. (2023). A Survey of Large Language Models for Healthcare: from Data, Technology, and Applications to Accountability and Ethics. *Information Fusion*, *118*(November 2024), 102963. https://doi.org/10.1016/j.inffus.2025.102963

Hermann, E. (2022). Leveraging Artificial Intelligence in Marketing for Social Good—An

Ethical Perspective. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-021-04843-y

Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. In *Journal of Strategic Information Systems*. https://doi.org/10.1016/j.jsis.2023.101772

Hofer, M., Obraczka, D., Saeedi, A., Köpcke, H., & Rahm, E. (2024). Construction of Knowledge Graphs: Current State and Challenges. *Information*, *15*(8). https://doi.org/10.3390/info15080509

Holzinger, A., Plass, M., Kickmeier-Rust, M., Holzinger, K., Crişan, G. C., Pintea, C.-M., & Palade, V. (2019). Interactive machine learning: experimental evidence for the human in the algorithmic loop. *Applied Intelligence*. https://doi.org/10.1007/s10489-018-1361-5

Hoseini, S., Theissen-Lipp, J., & Quix, C. (2024). A survey on semantic data management as intersection of ontology-based data access, semantic modeling and data lakes. *Journal of Web Semantics*, *81*(February), 100819. https://doi.org/10.1016/j.websem.2024.100819

Hu, Y., Liu, D., Wang, Q., Yu, C., Ji, H., & Xiong, J. (2024). Automating Knowledge Discovery from Scientific Literature via LLMs: A Dual-Agent Approach with Progressive Ontology Prompting. *Reserach Gate*.

Huettemann, S., & Mueller, R. (2025). *Enhancing Ontologies with Large Language Models: A Semi-Automated Approach*.

Ibrahim, N., Aboulela, S., Ibrahim, A., & Kashef, R. (2024). A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): models, evaluation metrics, benchmarks, and challenges. *Discover Artificial Intelligence*, *4*(1), 76. https://doi.org/10.1007/s44163-024-00175-8

Ignatious, H. A., El-Sayed, H., Khan, M. A., & Mokhtar, B. M. (2023). Analyzing Factors Influencing Situation Awareness in Autonomous Vehicles—A Survey. In *Sensors*. https://doi.org/10.3390/s23084075

Ismail, A. S., Ali, K. N., & Iahad, N. A. (2017). A Review on BIM-based automated code compliance checking system. *International Conference on Research and Innovation in Information Systems, ICRIIS*. https://doi.org/10.1109/ICRIIS.2017.8002486

Ivanisenko, T. V., Demenkov, P. S., & Ivanisenko, V. A. (2024). An Accurate and Efficient Approach to Knowledge Extraction from Scientific Publications Using Structured Ontology Models, Graph Neural Networks, and Large Language Models. *International Journal of Molecular Sciences*, *25*(21), 11811. https://doi.org/10.3390/ijms252111811

Jagannathan, M., & Delhi, V. S. K. (2019). Litigation Proneness of Dispute Resolution Clauses in Construction Contracts. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *11*(3). https://doi.org/10.1061/(ASCE)LA.1943-4170.0000301

Jain, S., & Mishra, S. (2016). Knowledge Representation with Ontology Tools & Methodology. *International Journal of Computer Applications® (IJCA*.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*. https://doi.org/10.1016/j.bushor.2018.03.007

Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., & Yu, S. (2021). A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2021.03.035

Jiang, J., Zhou, K., Dong, Z., Ye, K., Zhao, W. X., & Wen, J. R. (2023). StructGPT: A General Framework for Large Language Model to Reason over Structured Data.

EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings. https://doi.org/10.18653/v1/2023.emnlp-main.574

Jin, C., Chen, W., Cao, Y., Xu, Z., Tan, Z., Zhang, X., Deng, L., Zheng, C., Zhou, J., Shi, H., & Feng, J. (2020). Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nature Communications*. https://doi.org/10.1038/s41467-020-18685-1

Joachimiak, M. P., Miller, M. A., Caufield, J. H., Ly, R., Harris, N. L., Tritt, A., Mungall, C. J., & Bouchard, K. E. (2024). The Artificial Intelligence Ontology : LLM-assisted construction of AI concept hierarchies. *ArXiv Preprint*, 12.

Jofre, S. (2011). Strategic Management: The theory and practice of strategy in (business) organizations. DTU Management. *DTU Management*.

Jurisica, I., Mylopoulos, J., & Yu, E. (1999). Using Ontologies for Knowledge Management: An Information Systems Perspective. *Knowledge and Information Systems*, *6*. https://doi.org/10.1007/s10115-003-0135-4

Kaverinskiy, V., Palagin, O., & Litvin, A. (2024). Development and Testing of a Large Language Models Prompt for Natural Language Phrases Synthesis from Ontological Semantic Structures. *CEUR Workshop Proceedings*, *3806*, 253–264.

Keloth, V. K., He, Z., Chen, Y., & Geller, J. (2018). Leveraging Horizontal Density Differences between Ontologies to Identify Missing Child Concepts. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2018*, 644–653.

Khalef, R., & El-adaway, I. H. (2021). Automated Identification of Substantial Changes in Construction Projects of Airport Improvement Program: Machine Learning and Natural Language Processing Comparative Analysis. *Journal of Management in Engineering*, *37*(6). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000959

Khare, A., Ojala, A., & Baber, W. W. (2022). Sustainable International Business Models in a Digitally Transforming World. In *Sustainable International Business Models in a Digitally Transforming World*. https://doi.org/10.4324/9781003195986

Khyani, D., B S, S., Niveditha, N., M., D., & Y M, D. (2021). An Interpretation of Lemmatization and Stemming in Natural Language Processing. *Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology*, *22*, 350–357.

Kinney, M., Anastasiadou, M., Naranjo-Zolotov, M., & Santos, V. (2024). Expectation management in AI: A framework for understanding stakeholder trust and acceptance of artificial intelligence systems. *Heliyon*, *10*(7), e28562. https://doi.org/10.1016/j.heliyon.2024.e28562

Kirichenko, L., Radivilova, T., Bulakh, V., Zinchenko, P., & Saif Alghawli, A. (2020). Two approaches to machine learning classification of time series based on recurrence plots. *Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP 2020*. https://doi.org/10.1109/DSMP47368.2020.9204021

Kommineni, V. K., König-Ries, B., & Samuel, S. (2024). *From human experts to machines: An LLM supported approach to ontology and knowledge graph construction*.

Komninos, N., Bratsas, C., Kakderi, C., & Tsarchopoulos, P. (2016). Smart City Ontologies: Improving the effectiveness of smart city applications. *Journal of Smart Cities*. https://doi.org/10.18063/jsc.2015.01.001

Kostopoulou, O., Delaney, B. C., & Munro, C. W. (2008). Diagnostic difficulty and error in primary care--a systematic review. *Family Practice*, *25*(6), 400–413. https://doi.org/10.1093/fampra/cmn071

Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2021.102466

Kumar, S., Datta, S., Member, S., & Singh, V. (2024). *Applications , Challenges , and Future Directions of Human-in-the-Loop Learning*. *May*, 75735–75760.

Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ili, S., Hesslow, D., Castagné, R., Sasha Luccioni, A., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S., Webson, A., Sasanka Ammanamanchi, P., Wang, T., Sagot, B., Muennighoff, N., Villanova del Moral, A., … Belkada, Y. (2022). BLOOM: A 176B-Parameter Open-Access Multilingual Language Model Major Contributors Prompt Engineering Architecture and Objective Engineering Evaluation and Interpretability Broader Impacts. *ArXiv*.

Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, J. Y., Yoon, Y. G., Oh, T. K., Park, S., & Ryu, S. Il. (2020). A study on data pre-processing and accident prediction modelling for occupational accident analysis in the construction industry. *Applied Sciences (Switzerland)*, *10*(21), 1–23. https://doi.org/10.3390/app10217949

Lee, J., Yi, J.-S., & Son, J. (2019). Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP. *Journal of Computing in Civil Engineering*, *33*(3). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000807

Lepri, B., Oliver, N., & Pentland, A. (2021). Ethical machines: The human-centric use of artificial intelligence. In *iScience*. https://doi.org/10.1016/j.isci.2021.102249

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2020.acl-main.703

Li, G. (2017). Human-in-the-loop data integration. *Proceedings of the VLDB Endowment*, *10*(12), 2006–2017. https://doi.org/10.14778/3137765.3137833

Li, H., Yang, R., Xu, S., Xiao, Y., & Zhao, H. (2024). Intelligent Checking Method for Construction Schemes via Fusion of Knowledge Graph and Large Language Models. *Buildings*, *14*(8), 1–22. https://doi.org/10.3390/buildings14082502

Li, J., & Klinger, R. (2024). *Large Language Models with a Human in the Loop*. *1*(c).

Li, J., Yang, J., Hertzmann, A., Zhang, J., & Xu, T. (2021). LayoutGAN: Synthesizing Graphic Layouts with Vector-Wireframe Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2019.2963663

Li, S., Liu, C. H., Lin, Q., Wen, Q., Su, L., Huang, G., & Ding, Z. (2021). Deep Residual Correction Network for Partial Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2020.2964173

Li, X., Cui, M., Li, J., Bai, R., Lu, Z., & Aickelin, U. (2021). A hybrid medical text classification framework: Integrating attentive rule construction and neural network. *Neurocomputing*, *443*, 345–355. https://doi.org/10.1016/j.neucom.2021.02.069

Li, Z., & Ramani, K. (2007). Ontology-based design information extraction and retrieval. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*. https://doi.org/10.1017/S0890060407070199

Lin, Y., Pintea, S. L., & van Gemert, J. C. (2020). Deep Hough-Transform Line Priors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-58542-6_20

Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., Chowdhury, T., Li, Y., Cui, H., Zhao, T., Panalkar, A., Cheng, W., Wang, H., Liu, Y., Chen, Z., Chen, H., White, C., Gu, Q., Yang, C., & Zhao, L. (2023). *Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey*. https://doi.org/10.48550/arXiv.2305.18703

Lippolis, A. S., Ceriani, M., Zuppiroli, S., & Nuzzolese, A. G. (2025). Ontogenia: Ontology Generation with Metacognitive Prompting in Large Language Models. In A. Meroño Peñuela, O. Corcho, P. Groth, E. Simperl, V. Tamma, A. G. Nuzzolese, M. Poveda-Villalón, M. Sabou, V. Presutti, I. Celino, A. Revenko, J. Raad, B. Sartini, & P. Lisena (Eds.), *The Semantic Web: ESWC 2024 Satellite Events* (pp. 259–265). Springer Nature Switzerland.

Liu, B., Guo, W., Niu, D., Luo, J., Wang, C., Wen, Z., & Xu, Y. (2020). GIANT: Scalable Creation of a Web-scale Ontology. *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 393–409. https://doi.org/10.1145/3318464.3386145

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*. https://doi.org/10.1016/S2589-7500(19)30123-2

Lo, A., Jiang, A. Q., Li, W., & Jamnik, M. (2024). *End-to-End Ontology Learning with Large Language Models*. *NeurIPS*.

Lourdusamy, R., & Abraham, S. (2018). A Survey on Text Pre-processing Techniques and Tools. *International Journal of Computer Sciences and Engineering*, *06*(03), 148–157. https://doi.org/10.26438/ijcse/v6si3.148157

Loureiro, S. M. C., Guerreiro, J., & Tussyadiah, I. (2021). Artificial intelligence in business: State of the art and future research agenda. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2020.11.001

Luo, L., Li, Y. F., Haffari, G., & Pan, S. (2024). Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning. *12th International Conference on Learning Representations, ICLR 2024*, 1–24.

Ma, C., Molnár, B., & Benczúr, A. (2021). A semi-automatic semantic consistency-checking method for learning ontology from relational database. *Information (Switzerland)*, *12*(5). https://doi.org/10.3390/info12050188

Madubuko, C. C., Dip, G. C., Madubuko, C. C., & Chitsungo, C. (2024). *Applying Systems Theory to Ethical AI Development: Mitigating Unintended Consequences through Feedback Loop Analysis*. *09*(06), 1–31.

Mahadevkar, S. V., Patil, S., Kotecha, K., Soong, L. W., & Choudhury, T. (2024). Exploring AI-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-024-00948-z

Mahfouz, T., Kandil, A., & Davlyatov, S. (2018). Identification of latent legal knowledge in differing site condition (DSC) litigations. *Automation in Construction*, *94*, 104–111. https://doi.org/10.1016/j.autcon.2018.06.011

Makin, A. (2024a). *Ontology-Driven Knowledge Management Systems Enhanced by Large Language Models*. https://doi.org/10.13140/RG.2.2.24648.23043

Makin, A. (2024b). *Ontology-Driven Knowledge Management Systems Enhanced by Large Ontology-Driven Knowledge Management Systems Enhanced by Large Language Models AI and Ethics MSc. October*, 0–24. https://doi.org/10.13140/RG.2.2.32979.18728

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0194889

Mandal, S., & Connor, N. E. O. (2024). *LLMasMMKG : LLM Assisted Synthetic Multi-Modal Knowledge Graph Creation For Smart City Cognitive Digital Twins*.

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S., & Vaidya, P. (2023). Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. *Information Development*. https://doi.org/10.1177/02666669231200628

Mateiu, P., & Groza, A. (2023). Ontology engineering with Large Language Models. *Proceedings - 2023 25th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2023*, 226–229. https://doi.org/10.1109/SYNASC61333.2023.00038

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*.

McDaniel, M., & Storey, V. C. (2020). Evaluating Domain Ontologies. *ACM Computing Surveys*, *52*(4), 1–44. https://doi.org/10.1145/3329124

McKinsey Global Institute. (2016). The Age of Analytics: Competing in a Data-Driven World, Executive Summary. *McKinsey & Company, Brussels*.

Memarian, B., & Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100053. https://doi.org/10.1016/j.chbah.2024.100053

Memariani, A., Glauer, M., Neuhaus, F., Mossakowski, T., & Hastings, J. (2021). Automated and Explainable Ontology Extension Based on Deep Learning: A Case Study in the Chemical Domain. *CEUR Workshop Proceedings*.

Michie, S., Thomas, J., Johnston, M., Aonghusa, P. Mac, Shawe-Taylor, J., Kelly, M. P., Deleris, L. A., Finnerty, A. N., Marques, M. M., Norris, E., O'Mara-Eves, A., & West, R. (2017). The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implementation Science*, *12*(1), 121. https://doi.org/10.1186/s13012-017-0641-5

MIT Sloan Management Review. (2023). *Ontology-Driven AI: The Hidden Framework of Enterprise Success*.

Mollick, E., Mollick, L., Bach, N., Ciccarelli, L., Przystanski, B., & Ravipinto Generative Lab at Wharton, D. A. (2024). *Ai Agents and Education: Simulated Practice At Scale*.

Morandini, S., Fraboni, F., De Angelis, M., Puzzo, G., Giusino, D., & Pietrantoni, L. (2023). THE IMPACT OF ARTIFICIAL INTELLIGENCE ON WORKERS' SKILLS: UPSKILLING AND RESKILLING IN ORGANISATIONS. *Informing Science*. https://doi.org/10.28945/5078

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-022-10246-w

Mpofu, F. y. (2023). The application of Artificial Intelligence in external auditing and its implications on audit quality? A review of the ongoing debates. *International Journal* *of Research in Business and Social Science (2147- 4478)*, *12*(9), 496–512. https://doi.org/10.20525/ijrbs.v12i9.2737

Mulayim, O. B., Paul, L., Pritoni, M., Prakash, A. K., Sudarshan, M., & Fierro, G. (2024). Large Language Models for the Creation and Use of Semantic Ontologies in Buildings. *BuildSys 2024 - Proceedings of the 2024 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, *1*, 312–317. https://doi.org/10.1145/3671127.3698792

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). *A Comprehensive Overview of Large Language Models*.

Nawari, N. (2012). The challenge of computerizing building codes in a BIM environment. *Congress on Computing in Civil Engineering, Proceedings*. https://doi.org/10.1061/9780784412343.0036

Neuhaus, F. (2023). Ontologies in the era of large language models – a perspective. *Applied Ontology*, *18*(4), 399–407. https://doi.org/10.3233/AO-230072

Newton, J. (2018). Evolutionary game theory: A renaissance. In *Games*. https://doi.org/10.3390/g9020031

Nguyen, T. M., & Malik, A. (2022). A Two-Wave Cross-Lagged Study on AI Service Quality: The Moderating Effects of the Job Level and Job Role. *British Journal of Management*. https://doi.org/10.1111/1467-8551.12540

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, *104*(1), 11–33. https://doi.org/10.1109/JPROC.2015.2483592

Nie, J., Hou, X., Song, W., Wang, X., Zhang, X., Jin, X., Zhang, S., & Shi, J. (2024). *Knowledge Graph Efficient Construction : Embedding Chain-of-Thought into LLMs*.

Nobani, N. (2022). *Empowering XAI and LMI with Human-in-the-loop*. 1–182.

Norgeot, B., Glicksberg, B. S., & Butte, A. J. (2019). A call for deep-learning healthcare. In *Nature Medicine*. https://doi.org/10.1038/s41591-018-0320-3

Nowrozy, R., Ahmed, K., & Wang, H. (2024). *GPT, Ontology, and CAABAC: A Tripartite Personalized Access Control Model Anchored by Compliance, Context and Attribute*. https://doi.org/10.1371/journal.pone.0310553

Olga Perera, J. L. (2024). EXPLORING LARGE LANGUAGE MODELS FOR ONTOLOGY LEARNING. *Issues In Information Systems*, *25*(04). https://doi.org/10.48009/4_iis_2024_124

Osman, M. A., Mohd Noah, S. A., & Saad, S. (2022). Ontology-Based Knowledge Management Tools for Knowledge Sharing in Organization-A Review. In *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3163758

Ozkaya, I. (2023). Application of Large Language Models to Software Engineering Tasks: Opportunities, Risks, and Implications. *IEEE Software*, *40*, 4–8. https://doi.org/10.1109/MS.2023.3248401

Padiu, B., Iacob, R., Rebedea, T., & Dascalu, M. (2024). To What Extent Have LLMs Reshaped the Legal Domain So Far? A Scoping Literature Review. *Information (Switzerland)*, *15*(11), 1–25. https://doi.org/10.3390/info15110662

Palagin, O., Kaverinsky, V., Litvin, A., & Malakhov, K. (2023). Ontology-Driven Structured Prompts for ChatGPT Meta-Learning. *International Journal of Computing*, *22*(2), 170–183. https://doi.org/10.47839/ijc.22.2.3086

Palma, C. (2024). HILDEGARD: Human-in-the-Loop Data Extraction and Graphically Augmented Relation Discovery. *Journal of Computational and Cognitive Engineering*, *00*(March), 1–19. https://doi.org/10.47852/bonviewJCCE42022924

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and*

Data Engineering. https://doi.org/10.1109/TKDE.2024.3352100

Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2021.100838

Patil, R., & Gudivada, V. (2024). A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs). In *Applied Sciences (Switzerland)* (Vol. 14, Issue 5). https://doi.org/10.3390/app14052074

Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge Graphs: Opportunities and Challenges. In *Artificial Intelligence Review* (Vol. 56, Issue 11). Springer Netherlands. https://doi.org/10.1007/s10462-023-10465-9

Pereira, V., Hadjielias, E., Christofi, M., & Vrontis, D. (2023). A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2021.100857

Peter C. Verhoef, Thijs Broekhuizen, Yakov Bart, Abhi Bhattacharya, John Qi Dong, Nicolai Fabian, M. H. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, *122*, 889–901. https://doi.org/10.1016/j.jbusres.2019.09.022

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*. https://doi.org/10.18653/v1/n18-1202

Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. In *Harvard Business Review*.

Pospelov, D. (2023). *"Ontologies in the Digital Age: Advantages, Limitations, and Alternative Development Paths."*

Prikshat, V., Malik, A., & Budhwar, P. (2023). AI-augmented HRM: Antecedents, assimilation and multilevel consequences. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2021.100860

Radain, D. (2022). *The Use of Semantic Web Technologies in Business: Comparative Study*.

Radford, A. (2018). Improving Language Understanding by Generative Pre-Training. *Homology, Homotopy and Applications*, *9*(1), 399–438.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*.

Rameezdeen, R., & Rodrigo, A. (2014). Modifications to standard forms of contract: the impact on readability. *Construction Economics and Building*, *14*(2), 31–40. https://doi.org/10.5130/AJCEB.v14i2.3778

Rane, N., Choudhary, S., & Rane, J. (2024). A New Era of Automation in the Construction Industry: Implementing Leading-Edge Generative Artificial Intelligence, such as ChatGPT or Bard. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4681676

Reales, D., Manrique, R., & Grévisse, C. (2024). Core Concept Identification in Educational Resources via Knowledge Graphs and Large Language Models. *SN Computer Science*, *5*(8), 1029. https://doi.org/10.1007/s42979-024-03341-y

Regino, A. G., & Cesar, J. (2025). *Can LLMs be Knowledge Graph Curators for Validating Triple Insertions ? Iccl*, 87–99.

Ryabinin, K., Chuprina, S., & Belousov, K. (2019). Ontology-driven automation of IoT-based human-machine interfaces development. *Lecture Notes in Computer Science*

(Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-22750-0_9

Saeedizade, M. J., & Blomqvist, E. (2024a). Navigating Ontology Development with Large Language Models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *14664 LNCS*, 143–161. https://doi.org/10.1007/978-3-031-60626-7_8

Saeedizade, M. J., & Blomqvist, E. (2024b). Navigating Ontology Development with Large Language Models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 14664 LNCS* (pp. 143–161). https://doi.org/10.1007/978-3-031-60626-7_8

Saki Norouzi, S., Barua, A., Christou, A., Gautam, N., Eells, A., Hitzler, P., & Shimizu, C. (2024). *Ontology Population using LLMs*. https://doi.org/10.48550/arXiv.2411.01612

Salama, D. M., & El-Gohary, N. M. (2016). Semantic Text Classification for Supporting Automated Compliance Checking in Construction. *Journal of Computing in Civil Engineering*, *30*(1). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000301

Saseendran, A., Bigelow, B. F., Rybkowski, Z. K., & Jourdan, D. E. (2020). Disputes in Construction: Evaluation of Contractual Effects of ConsensusDOCS. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *12*(2). https://doi.org/10.1061/(ASCE)LA.1943-4170.0000377

Schadd, F. C., & Roos, N. (2015). Word-Sense Disambiguation for Ontology Mapping: Concept Disambiguation using Virtual Documents and Information Retrieval Techniques. *Journal on Data Semantics*, *4*(3), 167–186. https://doi.org/10.1007/s13740-014-0045-5

Schilling-Wilhelmi, M., Ríos-García, M., Shabih, S., Gil, M. V., Miret, S., Koch, C. T., Márquez, J. A., & Jablonka, K. M. (2024). *From Text to Insight: Large Language Models for Materials Science Data Extraction*. 1–51.

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. In *Neural Networks*. https://doi.org/10.1016/j.neunet.2014.09.003

Selma Wanna Fabian Parra, R. V. K. K., & Pryor, M. (2024). Unlocking underrepresented use-cases for large language model-driven human-robot task planning. *Advanced Robotics*, *38*(18), 1335–1348. https://doi.org/10.1080/01691864.2024.2366974

Sen, P., Mavadia, S., & Saffari, A. (2023). Knowledge Graph-augmented Language Models for Complex Question Answering. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2023.nlrse-1.1

Shen, H. T., Zhu, X., Zhang, Z., Wang, S. H., Chen, Y., Xu, X., & Shao, J. (2021). Heterogeneous data fusion for predicting mild cognitive impairment conversion. *Information Fusion*. https://doi.org/10.1016/j.inffus.2020.08.023

Shimizu, C., Hammar, K., & Hitzler, P. (2023). Modular ontology modeling. *Semantic Web*. https://doi.org/10.3233/SW-222886

Shimizu, C., Hirt, Q., & Hitzler, P. (2019). MODL: A modular ontology design library. *CEUR Workshop Proceedings*, *2459*, 47–58.

Shrestha, Y. R., Krishna, V., & von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2020.09.068

Siemens, G., Marmolejo-Ramos, F., Gabriel, F., Medeiros, K., Marrone, R., Joksimovic, S., & de Laat, M. (2022). Human and artificial cognition. *Computers and Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2022.100107

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Van Den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*. https://doi.org/10.1038/nature24270

Simmons, M., Singhal, A., & Lu, Z. (2017). Text Mining for Precision Medicine: Bringing structure to EHRs and biomedical literature to understand genes and health Michael. *Physiology & Behavior*, *176*(1), 100–106. https://doi.org/10.1177/0022146515594631.Marriage

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Snijder, L. L., Smit, Q. T. S., & De Boer, M. H. T. (2024). Advancing Ontology Alignment in the Labor Market: Combining Large Language Models with Domain Knowledge. *Proceedings of the AAAI Symposium Series*, *3*(1), 253–262. https://doi.org/10.1609/aaaiss.v3i1.31208

Solanki, A. (2024). Advancements in Artificial Intelligence: A Comprehensive Review and Future Prospects. *INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY*, *2*, 53–64.

Solihin, W., & Eastman, C. (2015). Classification of rules for automated BIM rule checking development. *Automation in Construction*. https://doi.org/10.1016/j.autcon.2015.03.003

Song, J., & Yoon, S. (2024). Ontology-assisted GPT-based building performance simulation and assessment: Implementation of multizone airflow simulation. *Energy and Buildings*, *325*, 114983. https://doi.org/https://doi.org/10.1016/j.enbuild.2024.114983

Song, S. (2017). Competition law and interoperability in cloud computing. *Computer Law* *and Security Review*. https://doi.org/10.1016/j.clsr.2017.05.005

Storey, V. C. (2019). Data management in the era of digitalization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-33223-5_1

Takerngsaksiri, W., Pasuksmit, J., Thongtanunam, P., Tantithamthavorn, C., Zhang, R., Jiang, F., Li, J., Cook, E., Chen, K., & Wu, M. (2024). *Human-In-the-Loop Software Development Agents*. *2*, 1–11.

Tan, X., Hammad, A., & Fazio, P. (2010). Automated Code Compliance Checking for Building Envelope Design. *Journal of Computing in Civil Engineering*. https://doi.org/10.1061/(asce)0887-3801(2010)24:2(203)

Thompson, N., Greenewald, K., Lee, K., & Manso, G. F. (2023). *The Computational Limits of Deep Learning*. https://doi.org/10.21428/bf6fb269.1f033948

Thuan, N. H., Antunes, P., & Johnstone, D. (2018). A Decision Tool for Business Process Crowdsourcing: Ontology, Design, and Evaluation. *Group Decision and Negotiation*, *27*(2), 285–312. https://doi.org/10.1007/s10726-018-9557-y

Tiddi, I., & Schlobach, S. (2022). Knowledge graphs as tools for explainable machine learning: A survey. *Artificial Intelligence*, *302*, 103627. https://doi.org/10.1016/j.artint.2021.103627

Tiwari, R. G., Husain, M., Srivastava, V., & Agrawal, A. (2011). Web personalization by assimilating usage data and semantics expressed in ontology terms. *International Conference and Workshop on Emerging Trends in Technology 2011, ICWET 2011 - Conference Proceedings*. https://doi.org/10.1145/1980022.1980133

Tocchetti, A., & Brambilla, M. (2022). The Role of Human Knowledge in Explainable AI. *Data*, *7*(7). https://doi.org/10.3390/data7070093

Trujillo, J. (2023). A PHENOMENOLOGICAL REPLY TO GARY MARCUS AND

ERNEST DAVIS'S REBOOTING AI: BUILDING ARTIFICIAL INTELLIGENCE WE CAN TRUST: CONTRIBUTING TO THE CREATION OF GENERAL ARTIFICIAL INTELLIGENCE. *Creativity Studies*. https://doi.org/10.3846/cs.2023.16193

Tsaneva, S., Dessì, D., Osborne, F., & Sabou, M. (2024). Enhancing Scientific Knowledge Graph Generation Pipelines with LLMs and Human-in-the-Loop. *Sci-K'24: 4th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment, November 2024, Baltimore, MD, USA, Co-Located with The 23rd International Semantic Web Conference, ISWC 2024*, *8418*, 0–9.

Tsaneva, S., & Sabou, M. (2024). Enhancing Human-in-the-Loop Ontology Curation Results through Task Design. *Journal of Data and Information Quality*, *16*(1), 1–25. https://doi.org/10.1145/3626960

Tudorache, T. (2020). Ontology engineering: Current state, challenges, and future directions. *Semantic Web*, *11*(1), 125–138. https://doi.org/10.3233/SW-190382

Tupayachi, J., Xu, H., Omitaomu, O. A., Camur, M. C., Sharmin, A., & Li, X. (2024). Towards Next-Generation Urban Decision Support Systems through AI-Powered Construction of Scientific Ontology Using Large Language Models—A Case in Optimizing Intermodal Freight Transportation. *Smart Cities*, *7*(5), 2392–2421. https://doi.org/10.3390/smartcities7050094

Ulansky, V., & Raza, A. (2021). Generalization of minimax and maximin criteria in a game against nature for the case of a partial a priori uncertainty. *Heliyon*. https://doi.org/10.1016/j.heliyon.2021.e07498

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review*. https://doi.org/10.1017/s0269888900007797

Val-Calvo, M., Egaña Aranguren, M., Mulero-Hernández, J., Almagro-Hernández, G., Deshmukh, P., Bernabé-Díaz, J. A., Espinoza-Arias, P., Sánchez-Fernández, J. L., Mueller, J., & Fernández-Breis, J. T. (2025). OntoGenix: Leveraging Large Language Models for enhanced ontology engineering from datasets. *Information Processing & Management*, *62*(3), 104042. https://doi.org/https://doi.org/10.1016/j.ipm.2024.104042

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Verma, S., Bhatia, R., Harit, S., & Batish, S. (2023). Scholarly knowledge graphs through structuring scholarly communication: a review. In *Complex and Intelligent Systems*. https://doi.org/10.1007/s40747-022-00806-6

Walker, J., Koutsiana, E., Nwachukwu, M., Meroño Peñuela, A., & Simperl, E. (2024). The Promise and Challenge of Large Language Models for Knowledge Engineering: Insights from a Hackathon. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613905.3650844

Walsh, K. P. (2017). Identifying and Mitigating the Risks Created by Problematic Clauses in Construction Contracts. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *9*(3). https://doi.org/10.1061/(ASCE)LA.1943-4170.0000225

Wang, P., Karigiannis, J., & Gao, R. X. (2024). Ontology-integrated tuning of large language model for intelligent maintenance. *CIRP Annals*, *73*(1), 361–364. https://doi.org/https://doi.org/10.1016/j.cirp.2024.04.012

Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., & Gao, J. (2020). Weak supervision for fake news detection via reinforcement learning. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v34i01.5389

Wang, Z., Joshi, S., Savel'Ev, S., Song, W., Midya, R., Li, Y., Rao, M., Yan, P., Asapu, S., Zhuo, Y., Jiang, H., Lin, P., Li, C., Yoon, J. H., Upadhyay, N. K., Zhang, J., Hu, M., Strachan, J. P., Barnell, M., … Yang, J. J. (2018). Fully memristive neural networks for pattern classification with unsupervised learning. *Nature Electronics*. https://doi.org/10.1038/s41928-018-0023-2

Washington, P. Y. (2022). *MACHINE LEARNING AND CROWDSOURCING FOR DIGITAL BEHAVIORAL PHENOTYPING*. *April*.

Wen, Y., Wang, Z., & Sun, J. (2023). *MindMap: Knowledge Graph Prompting Sparks Graph of Thoughts in Large Language Models*.

Wolniak. (2023). Functioning of real-time analytics in business. *Scientific Papers of Silesian University of Technology Organization and Management Series*, *2023*(172). https://doi.org/10.29119/1641-3466.2023.172.40

Wrick Talukdar, & Anjanava Biswas. (2023). Improving Large Language Model (LLM) fidelity through context-aware grounding: A systematic approach to reliability and veracity. *World Journal of Advanced Engineering Technology and Sciences*, *10*(2), 283–296. https://doi.org/10.30574/wjaets.2023.10.2.0317

Wu, X., & Lode, M. (2020). Language Models are Unsupervised Multitask Learners (Summarization). *OpenAI Blog*.

Wu, Y., Hu, N., Bi, S., Qi, G., Ren, J., Xie, A., & Song, W. (2023). *Retrieve-Rewrite-Answer: A KG-to-Text Enhanced LLMs Framework for Knowledge Graph Question Answering*.

Xiang, S., Zhen, C., Peng, J., Zhang, L., & Pu, Z. (2023). Application of Artificial Intelligence Technology in Power System Stability Assessment. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2023.11.011

Xie, J., Sun, L., & Zhao, Y. F. (2024). On the Data Quality and Imbalance in Machine

Learning-based Design and Manufacturing—A Systematic Review. *Engineering*, *xxxx*. https://doi.org/10.1016/j.eng.2024.04.024

Xin, D. D., Ma, L. L., Liu, J. J., Macke, S. S., Song, S. S., & Parameswaran, A. A. (2018). Accelerating Human-in-the-loop Machine Learning: Challenges and opportunities. *Proceedings of the 2nd Workshop on Data Management for End-To-End Machine Learning, DEEM 2018 - In Conjunction with the 2018 ACM SIGMOD/PODS Conference*. https://doi.org/10.1145/3209889.3209896

Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is inevitable: An innate limitation of large language models. *ArXiv Preprint ArXiv:2401.11817*.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. *NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*. https://doi.org/10.18653/v1/2021.naacl-main.41

Yan, H., Liu, Y., Jin, L., & Bai, X. (2023). The development, application, and future of LLM similar to ChatGPT. *Journal of Image and Graphics*, *28*(9), 2749–2762. https://doi.org/10.11834/jig.230536

Yang, H., Xiao, L., Zhu, R., Liu, Z., & Chen, J. (2024). *An LLM supported approach to ontology and knowledge graph construction*. 5240–5246. https://doi.org/10.1109/BIBM62325.2024.10822222

Yeh, C., Moritz, D., & Hohman, F. (2024). Exploring Empty Spaces : Human-in-the-Loop Data Augmentation. In *Proceedings of (ArXiv)* (Vol. 1, Issue 1). Association for Computing Machinery.

Yihang Zhao, N. V. A. and K., & Aryan. (2024). *Using Large Language Models for OntoClean-based Ontology Refinement*. *August 2023*, 1–10.

Yin, Y., Wang, B., Zuo, H., & Childs, P. (2024, August). Effects of Different Human-In-The-Loop Approaches on Human-AI Co-Design: A Comparison Between Human-Learning HITL Approach and Machine-Learning HITL Approach. *Volume 2A: 44th Computers and Information in Engineering Conference (CIE)*. https://doi.org/10.1115/DETC2024-143295

Yuan, M., Bao, P., Yuan, J., Shen, Y., Chen, Z., Xie, Y., Zhao, J., Li, Q., Chen, Y., Zhang, L., Shen, L., & Dong, B. (2024). Large language models illuminate a progressive pathway to artificial intelligent healthcare assistant. *Medicine Plus*, *1*(2), 100030. https://doi.org/10.1016/j.medp.2024.100030

Yun, H., Xu, J., Xiong, J., & Wei, M. (2011). A knowledge engineering approach to develop domain ontology. *International Journal of Distance Education Technologies*. https://doi.org/10.4018/jdet.2011010104

Zait, F., & Zarour, N. (2018). Addressing Lexical and Semantic Ambiguity in Natural Language Requirements. *2018 Fifth International Symposium on Innovation in Information and Communication Technology (ISIICT)*, 1–7. https://doi.org/10.1109/ISIICT.2018.8613726

Zengeya, T., & Vincent Fonou-Dombeu, J. (2024). A Review of State of the Art Deep Learning Models for Ontology Construction. *IEEE Access*, *12*(April), 82354–82383. https://doi.org/10.1109/ACCESS.2024.3406426

Zhang, B., Peñuela, A. M., & Simperl, E. (2023a). Towards Explainable Automatic Knowledge Graph Construction with Human-in-the- Loop. *Frontiers in Artificial Intelligence and Applications*. https://doi.org/10.3233/FAIA230091

Zhang, B., Peñuela, A., & Simperl, E. (2023b). *Towards Explainable Automatic Knowledge Graph Construction with Human-in-the-Loop*. https://doi.org/10.3233/FAIA230091

Zhang, X., Wang, S., Liu, J., & Tao, C. (2018). Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. *BMC Medical Informatics and Decision Making*. https://doi.org/10.1186/s12911-018-0631-9

Zhang, Z., Gu, Y., Han, X., Chen, S., Xiao, C., Sun, Z., Yao, Y., Qi, F., Guan, J., Ke, P., Cai, Y., Zeng, G., Tan, Z., Liu, Z., Huang, M., Han, W., Liu, Y., Zhu, X., & Sun, M. (2021). CPM-2: Large-scale cost-effective pre-trained language models. *AI Open*. https://doi.org/10.1016/j.aiopen.2021.12.003

Zhang, Z., Sjons, J., Rajwade, R., & Ab, R. (2023). *Text Normalization for Text-to-Speech*.

Zhao, S., Liu, Z., Lin, J., Zhu, J. Y., & Han, S. (2020). Differentiable augmentation for data-efficient GAN training. *Advances in Neural Information Processing Systems*.

Zheng, N. ning, Liu, Z. yi, Ren, P. ju, Ma, Y. qiang, Chen, S. tao, Yu, S. yu, Xue, J. ru, Chen, B. dong, & Wang, F. yue. (2017). Hybrid-augmented intelligence: collaboration and cognition. *Frontiers of Information Technology and Electronic Engineering*, *18*(2), 153–179. https://doi.org/10.1631/FITEE.1700053

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. *Advances in Neural Information Processing Systems*.

Zhou, P., & El-Gohary, N. (2016). Automated Extraction of Environmental Requirements from Contract Specifications. *... Conference on Computing in Civil and …*.

Zhuang, Y. ting, Wu, F., Chen, C., & Pan, Y. he. (2017). Challenges and opportunities: from big data to knowledge in AI 2.0. In *Frontiers of Information Technology and Electronic Engineering*. https://doi.org/10.1631/FITEE.1601883

Zirar, Araz, Ali, Imran, S., Islam, & Nazrul. (2023). Worker and workplace Artificial Intelligence (AI) coexistence: Emerging themes and research agenda. *Technovation*, *124*, 102747. https://doi.org/10.1016/j.technovation.2023.102747

Zou, Y., Kiviniemi, A., & Jones, S. W. (2017). Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Automation in Construction*, *80*, 66–76. https://doi.org/10.1016/j.autcon.2017.04.003

APPENDIX A:

```
import nltk #It means that we will import necessary modules from NLTK
(Natural Language Toolkit)
nltk.download('stopwords') #Download the stopwords files from the NLTK
library which are used in the text preprocessing to remove most common
words such as the, and, as, etc.
nltk.download('punkt_tab') #Download the ‘punkt_tab’ tokenizer models says
for sentence and word tokenization.
nltk.download('wordnet') #The wordnet dataset needs to be downloaded from
NLTK and is often used for lemmatization purposes; it is a lexical
database for English.

#Handle Missing Values
#Fill missing categorical values with mode
CategoricalColumns = dataOsha.select_dtypes(include=['object']).columns
for cl in CategoricalColumns:
    dataOsha[cl].fillna(dataOsha[cl].mode()[0], inplace=True)

#Fill missing numerical values with median
NumericalColumns = dataOsha.select_dtypes(include=['int64',
'float64']).columns
imputer = SimpleImputer(strategy='median')
dataOsha[NumericalColumns] =
imputer.fit_transform(dataOsha[NumericalColumns])

#Standardize Date Format
dataOsha['Event Date'] = pnds.to_datetime(dataOsha['Event Date'],
errors='coerce')

#Normalize Text Fields
stpwords = set(stopwords.words('english'))
lemat = WordNetLemmatizer()

def preprocess_text(txxt):

text_columns = ['Abstract Text', 'Event Description', 'Event Keywords']
for col in text_columns:
    dataOsha[col] = dataOsha[col].apply(preprocess_text)

#Encode Categorical Columns (Optional)
#Example: One-hot encoding for 'Degree of Injury'
```

```
dataOsha = pnds.get_dummies(dataOsha, columns=['Degree of Injury'],
drop_first=True)

#Save Preprocessed Data

# Most frequent keywords in 'Event Keywords'
from collections import Counter

import seaborn as saas

event_keywords = ' '.join(dataOsha['Event Keywords'])
event_keywords_count = Counter(event_keywords.split())
top_keywords = event_keywords_count.most_common(10)

# Bar plot for top 10 keywords
keywords, counts = zip(*top_keywords)

#Load your OSHO dataset (replace with your dataset's path)
filepath = "/content/preprocessed_osho_dataset.csv"
osha_data = pnds.read_csv(filepath)
osha_data = osha_data.head(100)

#Apply the function to a subset of the "Event Description" column
osha_data["Generated Ontology"] = osha_data["Event
Description"].apply(generate_ontology)

#Save the results to a JSON file
ontology_results = osha_data[["summary_nr", "Generated
Ontology"]].to_dict(orient="records")
output_file = "ontology_results.json"

cleaned_data = []
for record in ontology_data:
    try:
        ontology_json = json.loads(record["Generated
Ontology"].strip("```json").strip())
        cleaned_data.append({
            "summary_nr": record["summary_nr"],
            "Entities": ontology_json.get("Entities", []),
            "Relationships": ontology_json.get("Relationships", [])
        })
    except Exception as e:
```

```
        print(f"Error parsing ontology for summary_nr
{record['summary_nr']}: {e}")

# Convert to a DataFrame for easier manipulation
ontology_dataOsha = pnds.DataFrame(cleaned_data)
ontology_dataOsha.head()

from collections import Counter

# Analyze Entities
entity_counter = Counter()
for entities in ontology_dataOsha["Entities"]:
    entity_counter.update(entities)

# Analyze Relationships
relationship_counter = Counter()
for relationships in ontology_dataOsha["Relationships"]:
    relationship_counter.update(relationships)

# Top 20 entities
top_entities = entity_counter.most_common(20) #A list of 20 entities which
have the highest counts can also be extracted from the entity counter.
entities, counts = zip(*top_entities) #Unzip the list of tuples into two
separate lists: realization in the form of pursuing entities (names) and
counts (frequencies).

mlpt.figure(figsize=(10, 6)) #Generate a figure that will be used for the
bar plot and the figure should have the desired size of (width=10 inch and
height=6 inch).
# Enjoy seaborn’s barplot function to produce a horizontally oriented bar
figure
# x-axis represents the count that is frequency, y-axis represents
entities and ‘palette’ sets the bars color.
saas.barplot(x=list(counts), y=list(entities), palette="pastel")
mlpt.title("Top 20 Entities in Ontology") #Assign the title of the plot to
explain the visualization
mlpt.xlabel("Frequency") #Make a note on the x-axis telling what is
represented on the horizontal axis
mlpt.ylabel("Entities") #Put a label on the left vertical axis to show
what is measured on the given axis
mlpt.show() #Display the bar plot
```

```
# Top 20 relationships
top_relationships = relationship_counter.most_common(20) #Get the 20 most
frequent relationships and their frequencies from the relation counter.
relationships, counts = zip(*top_relationships) #Unzip the list of tuples
into two separate lists: events (names) and occurrences (frequencies).

mlpt.figure(figsize=(10, 6)) #To build the figure for the bar plot,
propose a specific size: width = 10 inches, height = 6 inches.
#To create the bar plot you should use seaborn’s barplot function but with
horizontal parameter set to True.
#Assign the x-axis to the count (frequency) aspects and the y-axis to be
the relationship spectrums of the samples chosen while applying pastel
color to the bars.
saas.barplot(x=list(counts), y=list(relationships), palette="pastel")
mlpt.title("Top 20 Relationships in Ontology") #In titles of the plot, set
the title of the plot to describe the visualization.
mlpt.xlabel("Frequency") #Describe briefly on what you would label your x-
axis if you were to interpret it generally.
mlpt.ylabel("Relationships") #Make sure to put a label at the y-axis so
that everyone that looks at it will know what the vertical axis is
displaying.
mlpt.show() #Display the bar plot.

def visualize_entity_relationships(ontology_dataOsha, target_entity,
top_n=50):
    """
    Visualizes relationships involving a specific entity.
    """
    G = nex.DiGraph()  # Directed graph
    for _, row in ontology_dataOsha.iterrows():
        for relationship in row["Relationships"]:
            if target_entity in relationship:
                try:
                    entity1, rel, entity2 = map(str.strip,
relationship.split("->"))
                    G.add_edge(entity1, entity2, label=rel)
                except ValueError:
                    continue

    # Plot the graph

import networkx as nex
```

```
# Function to visualize relationships
def visualize_relationships(relationships, top_n=50):
    G = nex.DiGraph()  # Directed graph for relationships
    for relationship in relationships[:top_n]:
        try:
            entity1, rel, entity2 = map(str.strip, relationship.split("-
>"))
            G.add_edge(entity1, entity2, label=rel)
        except ValueError:
            print(f"Skipping malformed relationship: {relationship}")

            # Parse the Generated Ontology JSON field
            generated_ontology = json.loads(record["Generated
Ontology"].strip("```json").strip())
            entities = generated_ontology.get("Entities", [])
            relationships = generated_ontology.get("Relationships", [])

            review_data.append({
                "summary_nr": record["summary_nr"],
                "Entities": ", ".join(entities),  # Combine entities into
a single string
                "Relationships": "; ".join(relationships)  # Combine
relationships into a single string
            })
        except Exception as e:
            print(f"Error processing record {record['summary_nr']}: {e}")

    # Check if review_data is populated
    if not review_data:
        print("No data to save. Please check the ontology file
structure.")
        return

    # Convert to DataFrame and save to CSV
    review_dataOsha = pnds.DataFrame(review_data)
    try:


# Normalize entities and relationships
def normalize_text(text):
    """Convert text to lowercase and remove extra spaces."""
```

```
    return text.lower().strip()

# Safely apply normalization to entities and relationships
def normalize_column(column, delimiter):
    """Normalize a column with a given delimiter."""
    def normalize_entry(entry):
        if isinstance(entry, str):  # Check if entry is a string
            return delimiter.join([normalize_text(item) for item in
entry.split(delimiter)])
        return ""  # Handle non-string entries (e.g., NaN)
    return column.apply(normalize_entry)

# Apply normalization to Entities and Relationships
dataOsha["Entities"] = normalize_column(dataOsha["Entities"], ",")
dataOsha["Relationships"] = normalize_column(dataOsha["Relationships"],
";")

print("After Normalization:")
dataOsha.head()

# Refine relationships
def refine_relationships(relationships):
    """Refine generic or ambiguous relationships."""
    refined = []
    for rel in relationships.split(";"):
        if "experienced" in rel:
            refined.append(rel.replace("experienced", "suffered"))
        elif "occurred on" in rel:
            refined.append(rel.replace("occurred on", "occurred from"))
        elif "hospitalized" in rel:
            refined.append(rel.replace("hospitalized", "hospitalized due
to"))
        elif "struck" in rel:
            refined.append(rel.replace("struck", "hit by"))
        else:
            refined.append(rel)
    return "; ".join(refined)

# Apply refinement
dataOsha["Relationships"] =
dataOsha["Relationships"].apply(refine_relationships)
```

```
print("After Refining Relationships:")
dataOsha.head()

# Combine related entities
def combine_related_entities(entities):
    """Combine related entities into compound entities."""
    entity_list = entities.split(", ")
    if "firework" in entity_list and "incident" in entity_list:
        entity_list.remove("firework")
        entity_list.remove("incident")
        entity_list.append("firework accident")
    return ", ".join(entity_list)

# Apply combination
dataOsha["Entities"] =
dataOsha["Entities"].apply(combine_related_entities)

print("After Combining Related Entities:")
dataOsha.head()

# Remove duplicate entities
def remove_duplicates(entities):
    """Remove duplicate entities."""
    entity_list = entities.split(", ")
    return ", ".join(sorted(set(entity_list)))

# Apply duplicate removal
dataOsha["Entities"] = dataOsha["Entities"].apply(remove_duplicates)

print("After Removing Redundancies:")
dataOsha.head()

# Save the refined data to a new CSV file
refined_filepath = "/content/drive/MyDrive/Ramanathan
Iyer/Files/refined_ontology_for_review.csv"
dataOsha.to_csv(refined_filepath, index=False)

print(f"Refined ontology data saved to {refined_filepath}")

# Load the datasets
automated_dataOsha = pnds.read_csv("/content/drive/MyDrive/Ramanathan
Iyer/Files/ontology_for_review.csv")  # Automated ontologies
```

```
human_reviewed_dataOsha = pnds.read_csv("/content/drive/MyDrive/Ramanathan
Iyer/Files/refined_ontology_for_review.csv")  # Human-reviewed ontologies

# Function to calculate evaluation metrics for relationships only
def evaluate_relationships(dataOsha, reference_dataOsha=None,
domain_relevance_rules=None):
    """
    Evaluates relationships between automated and human-reviewed datasets.
    matix:
    - Accuracy: Proportion of correct relationships compared to the
reference.
    - Completeness: Average number of relationships per record.
    - Relevance: Proportion of relevant relationships (requires domain
rules).
    - Consistency: Proportion of records with no duplicate relationships.
    """
    def normalize_list(items):
        """Normalize and clean relationship list."""
        return [item.strip().lower() for item in items]

    matix = {
        "accuracy_relationships": 0.0,
        "completeness_relationships": 0.0,
        "relevance_relationships": 0.0,
        "consistency_relationships": 0.0,
    }

    total_records = len(dataOsha)
    total_relationships = 0
    consistent_relationships = 0
    relevant_relationships_count = 0
    correct_relationships = 0

    for _, row in dataOsha.iterrows():
        # Handle missing or non-string values
        relationships = str(row["Relationships"]) if
isinstance(row["Relationships"], str) else ""
        relationships = normalize_list(relationships.split("; ")) if
relationships else []

        # Completeness
        total_relationships += len(relationships)
```

```
        # Consistency
        if len(relationships) == len(set(relationships)):  # No duplicate
relationships
            consistent_relationships += 1

        # Relevance
        if domain_relevance_rules:
            relevant_relationships_count += sum(1 for relationship in
relationships if any(term in relationship for term in
domain_relevance_rules))

        # Accuracy (compare with reference_dataOsha)
        if reference_dataOsha is not None:
            reference_row =
reference_dataOsha[reference_dataOsha["summary_nr"] == row["summary_nr"]]
            if not reference_row.empty:
                ref_relationships =
str(reference_row.iloc[0]["Relationships"]) if
isinstance(reference_row.iloc[0]["Relationships"], str) else ""
                ref_relationships =
set(normalize_list(ref_relationships.split("; "))) if ref_relationships
else set()

                # Calculate accuracy
                correct_relationships += len(set(relationships) &
ref_relationships)

    # Calculate metrics
    matix["accuracy_relationships"] = correct_relationships /
total_relationships if total_relationships > 0 else 0
    matix["completeness_relationships"] = total_relationships /
total_records if total_records > 0 else 0
    matix["relevance_relationships"] = relevant_relationships_count /
total_relationships if total_relationships > 0 else 0
    matix["consistency_relationships"] = consistent_relationships /
total_records if total_records > 0 else 0

    return matix

# Define domain-specific relevance rules (example terms)
```

```
domain_relevance_rules = ["employee", "fall", "injury", "machine", "roof",
"accident", "hospitalized"]

# Evaluate relationships in automated and human-reviewed datasets
print("Evaluating Relationships (Automated vs Human-Reviewed)...")
relationship_matix = evaluate_relationships(automated_dataOsha,
human_reviewed_dataOsha, domain_relevance_rules)

# Display metrics
print("\nRelationship metrics:")
for key, value in relationship_matix.items():
    print(f"{key}: {value:.2f}")
```