Towards Explainable Commonsense Reasoning: Semantic Rule Generation from Text using LLMs

Muhammad Raza Naqvi[1,*,†], Arkopaul Sarkar[2,‡], Antoine Zimmermann[1,†], Bernard Archimede[4,†], Linda Elmhadhbi[3,†] and Mohamed Hedi Karray[4,†]

[1]Mines Saint-Etienne, Univ Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS, F-42023 Saint-Etienne, France.

[2]Department of Physics, Georgetown University, 37th St NW,Washington, DC 20057, USA

[3]INSA Lyon, Université Lumière Lyon 2, Université Claude Bernard Lyon 1, Université Jean Monnet Saint-Etienne, DISP UR4570, Villeurbanne, France

[4]Laboratoire Génie de Production, Université de Technologie Tarbes Occitanie Pyrénées (UTTOP), 47 Av. d'Azereix, Tarbes, 65016, France.

Abstract

Commonsense knowledge (CSK) is critical for enhancing artificial intelligence (AI) systems by improving their understanding, reasoning, and interaction with the human world, particularly in planning and decision-making tasks. To be practically applicable, CSK must be expressed using the standard vocabulary of the target domain and must be available in sufficient quantity and specificity. Large Language Models (LLMs) have shown promise in efficiently curating domain-specific CSK in natural language statements. However, transforming these statements into formal semantic rules such as those written in the Semantic Web Rule Language (SWRL) or Datalog requires further processing and structured prompt engineering. These models also often fail to incorporate standard vocabularies such as those defined by ISO 21838 when generating such rules, limiting their interoperability and reuse. This paper addresses the interoperability challenges in capturing CSK and highlights the importance of standardized vocabularies for semantic integration. We propose a template-based prompt-engineering method combined with a predefined vocabulary-to-ontology mapping to guide LLMs in generating semantic rules from natural language CSK. Our findings reveal key limitations in the ability of LLMs to align output with standard ontologies. To address this, we propose a template-based prompt-engineering method combined with a predefined vocabulary-to-ontology mapping. Comparative evaluation shows that our approach improves consistency and enhances alignment with upper-level ontologies when expressing CSK as semantic rules.

Keywords

Common Sense Knowledge, Knowledge Engineering, Large Language Models, Manufacturing Common-Sense knowledge, Semantic Explainable AI,

1. Introduction

In the era of AI, CSK is a crucial component of AI systems that enables them to make rational and explainable decisions, much like humans do [1]. CSK is an essential element of today's AI-driven decision-making applications. When it comes to sophisticated tasks and interactions across different domains, it is pivotal for AI systems to be equipped with this type of knowledge.

XAI-KRKG@ECAI25: First International ECAI Workshop on eXplainable AI, Knowledge Representation and Knowledge Graphs, October 25–30, 2025, Bologna, Italy

*Corresponding author. razisyed4@gmail.com

This form of knowledge includes the implicit and frequently used understanding of the world that humans naturally possess [2]. Researchers in various domains are increasingly emphasizing the importance of acquiring and integrating appropriate domain-specific CSK [3, 4, 5].

using a large volume of domain-specific CSK improves the AI system's ability to make decisions effectively and in an explainable manner. It also enables the system to adapt appropriately to different scenarios [6, 7]. The emergence of LLMs has initiated a new era in which these models possess vast amounts of embedded knowledge across numerous domains. As a result of the extensive information they contain, LLMs can serve as a surface-level source of commonsense-like assertions across a wide range of domains; however, their reliability and semantic coherence as knowledge sources remain limited and often require external validation. The realization that these models are trained on massive textual datasets and reflect a broad spectrum of human knowledge makes them highly valuable for capturing CSK [8].

However, the transition from the unstructured, semantically ambiguous CSK generated by LLMs to structured, logically coherent semantic rules remains a significant challenge [9, 10, 11]. Semantic rules are crucial for explainability because they provide explicit, human-readable logic that governs system decisions, enabling users to trace why and how a particular output was produced. Unlike black-box models, semantic rules support transparent reasoning by linking inputs to conclusions through well-defined conditions grounded in domain knowledge. This clarity enables justified explanations, particularly in high-stakes contexts such as manufacturing or healthcare. This is particularly evident when using GPTs (Generative Pre-trained Transformers) for automatic knowledge engineering [12, 13]. While the evolution of LLMs raises questions about the extent to which such models might be integrated into various industries, businesses, and most importantly education it also brings forward critical issues such as ethics [14, 15], trustworthiness [16, 17], and adherence to Findable, Accessible, Interoperable, Reusable (FAIR) data principles [18, 19]. In the specific context of knowledge engineering, particularly ontology development, one might ask: Are we heading toward a future where LLMs automatically generate ontologies, potentially rendering human ontologists obsolete? [20].

We argue that ontology engineering and mapping extend far beyond mere linguistic tasks. While LLMs are proficient at tasks such as relation extraction and entity recognition, both of which support ontology engineering, true ontology development requires input from domain experts to define terms, structure hierarchical relationships, and provide formal representations grounded in logical inference. Furthermore, one of the keys to making ontologies FAIR and interoperable is aligning them with standard vocabularies, such as ISO-21838, which are closely linked to commonsense knowledge. Ontologies are consistently validated by domain experts and evolve, whereas the validation of information produced by LLMs remains an open issue [16, 17].

This paper proposes a methodology for generating semantic rules from CSK statements, guided by predefined mappings to standard ontologies. It argues for the continued importance of ontologies and the necessity of human involvement in the development process, particularly for capturing and incorporating CSK from LLMs. This approach enables the derivation of rule-based expressions such as First Order Logic (FOL) that can inform the creation of ontology classes and properties using standard vocabularies. When LLMs are prompted to generate NL statements based on CSK, these statements are then transformed into First-Order Logic (FOL) rules based on CSK patterns aligned with standard vocabulary. This rule-based method addresses key limitations of LLMs by producing formal, vocabulary-aligned rules that support the development of consistent, standardized, and semantically rich ontologies, in adherence to principles of formal logic.

The remainder of the paper is structured as follows: Section 2 provides a brief overview of how LLMs operate based on textual patterns, Section 3 discusses prompt engineering techniques, and examines the limitations of LLMs in knowledge engineering, along with the need for human involvement. Section 4 presents the proposed methodology based on CSK-driven semantic rules and predefined mappings to standard ontologies. Section 5 assesses the effectiveness and applicability of our method, also its limitations, and Section 6 demonstrates the applicability of the proposed methodology in the manufacturing domain, and lastly, Section 7 concludes the paper and outlines directions for future work.

2. Literature Review

Understanding the distinction between facts and knowledge is pivotal to addressing the challenge of ontology development using LLMs [21]. A fact is "a statement that can be proven to be true or false," whereas knowledge, in the context of ontologies and knowledge engineering, encompasses a broader understanding that includes the interpretation and inference of facts within a certain domain. Knowledge is not about truthfulness but involves the structured organization and representation of information that can be used to infer new insights [22]. CSK is a subset of knowledge that is considered to be universally true [23] and is crucial for the development of ontologies that accurately reflect real-world semantics [24]. When discussing CSK, the Cyc project aimed to develop an ontology containing common knowledge terms, facts, concepts, and rules. The project also focused on creating a system capable of communicating in English and learning from human interactions [25]. This goal has now been partially achieved by LLMs, which can interact with users in natural language and answer queries using different prompts, although without creating ontologies, a central aim of the Cyc project.

2.1. The Evolution and Impact of LLMs in AI

LLMs represent a significant era in technological innovation. They have not only reshaped the landscape of AI but also ushered in a new research era focused on Generative AI. The evolution of Natural Language Processing (NLP) has played a critical role in enabling machines to read, understand, and make sense of human language, alongside Machine Learning (ML) systems that facilitate the development of models capable of making predictions from data [26, 27, 28, 29]. As we explore LLMs further, it becomes evident that models like GPTs (Generative Pre-trained Transformers) signify a major leap in AI. LLMs are designed to understand, interact, and generate language at an unprecedented scale, owing to their access to massive text corpora from which they learn linguistic patterns and structures. This enables them to perform a wide range of language-based tasks with remarkable efficiency [30]. The capabilities of these models have sparked debates about whether LLMs are on par with humans in everyday tasks and the societal implications of their use [31]. Today, it is common to find examples showcasing LLMs as either impressively intelligent or inexplicably flawed, often referred to as hallucinations."Regardless, they demonstrate a notable ability to process and respond to human language, often requiring substantial background knowledge [32].

2.2. Challenges in Using LLMs for Ontology Development

Despite the substantial advancements achieved by LLMs, their application in ontology development presents unique challenges due to insufficient knowledge modeling and limited reasoning capabilities [33]. In knowledge engineering, ontologies represent a set of concepts within a domain and the relationships between them. They are essential for reasoning about entities and making inferences. Ontology development entails creating a standardized vocabulary and expressing formal semantics through axioms. The challenge lies in the design of LLMs: they excel in statistical pattern recognition but struggle with the deep semantic structures and formal logic required for ontology construction [34, 35]. LLMs lack the ability to grasp intricate and nuanced relationships and classifications, which are essential for accurate ontology development [36]. Although LLMs can comprehend and generate text using statistical patterns, they cannot understand the semantic relationships and logical structures necessary for creating meaningful ontologies [37, 38]. This limitation highlights the need for innovative approaches that meet the requirements of formal logic and semantic complexity [39].

3. Ontology Development Practices and the Role of Pre-Trained LLMs

Ontologies, especially reference ontologies, are developed to provide a standardized, controlled vocabulary for a specific domain. For example, the IOF Core Ontology encompasses notions common across multiple manufacturing domains, while top-level ontologies such as the Basic Formal Ontology (BFO)[1], Suggested Upper Merged Ontology (SUMO), and OpenCyc address more general conceptualizations [40].

An open-source reference ontology offers human- and machine-readable definitions of its vocabulary. A major use case for reference ontologies is enabling interoperability between datasets that use these standardized terms in their data or metadata.

LLMs can provide diverse information based on user prompts via natural language interactions. Numerous prompt engineering techniques are available, as summarized by Schmidt et al. [41]. Pre-trained LLMs such as OpenAI's ChatGPT series and Google's BERT have shown effectiveness in various NLP tasks. While their primary role is content generation and language translation, they are now also being explored for ontology creation due to several capabilities: (1) Semantic Understanding, (2) Entity Recognition, (3) Relation Extraction, and (4) Concept Generation. Trained on vast amounts of text, these models capture relationships such as synonymy, hyponymy, and hypernymy due to their semantic capabilities [42]. They also perform well on Named Entity Recognition (NER) tasks, identifying entities such as locations, organizations, and people, and can extract relationships between mentioned entities [43]. According to Grandi et al., LLMs can generate conceptual designs based on prompts [44].

[1]https://basic-formal-ontology.org

3.1. Limitations of LLMs in Knowledge Engineering

Despite these capabilities, LLMs exhibit several limitations that render them unsuitable as standalone tools for ontology development [44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57]: **(1)** Lack of explicit knowledge representation, **(2)** Lack of logical consistency, **(3)** Semantic ambiguity and inconsistent responses, **(4)** Domain specificity, **(5)** Data bias and incompleteness, **(6)** Limited multi-modal understanding, **(7)** Scalability issues.

3.2. Comparison with Existing Pre-Trained GPTs

LLMs generate responses based on patterns in training data, not on explicit semantic structures like those in ontologies. When generating definitions or relationships between terms, LLMs may produce inconsistent information due to inaccurate representations in their training data. A key issue is the inability of LLMs to generate ontologies aligned with formal ontological standards. While they can convert natural language into semantic rules (e.g., SPARQL, SWRL, or Datalog), they often invent predicates instead of reusing existing ones from standard ontologies. This behavior underscores the need for aligning LLM-generated content with established standards, such as Top-Level Ontologies (TLOs) like the Basic Formal Ontology (BFO) **[ISO/IEC 21838-2:2021]** and mid-level ontologies like the Industrial Ontologies Foundry (IOF)[2], and Machine Service Description Language (MSDL)[3]. For general-purpose vocabularies like FOAF (Friend of a Friend) or family trees, LLMs can effectively model text and align with standard ontologies, given the generality of linguistic terms and labels. However, modeling abstract concepts using vocabularies like BFO is more difficult. In our prompt-based use case with GPT-3.5[4], the model often failed to reuse the provided ontological terms and instead generated plausible-sounding but ontologically invalid relations, highlighting its dependence on linguistic surface patterns rather than formal alignment.

GPT-3.5 also faces input limitations due to the 4096-token cap, which prevents users from uploading large RDF or TTL files. In GPT-4.0[5], users can upload ontologies as files, improving results. Nevertheless, the model still fails to consistently reuse the provided ontological classes and relations.

4. Methodology

The proposed method[6] addresses the significant challenge of translating Common Sense Knowledge (CSK) into semantic rules. These rules are retrieved from a Large Language Model (LLM) as natural language (NL) statements and subsequently converted into ontology classes, subclasses, instances, or relationships, while ensuring alignment with standard vocabularies. The primary goal is to bridge the gap between the flexible nature of natural language and the structured, rigid requirements of ontologies necessary for effective reasoning and data integration.

[2] https://spec.industrialontologies.org/iof/
[3] https://labs.engineering.asu.edu/semantics/ontology-download/msdl-ontology/
[4] https://chat.openai.com/share/44a3d1d6-66b2-407d-b6db-509f158a30f9
[5] https://chat.openai.com/share/86565869-7a00-47e8-9067-d6359f61c32c
[6] https://github.com/MRNaqvi/Common-Sense-Knowledege-Driven-SemanticRule-Base-Ontology-Mapping

To achieve this, a rule-based mechanism identifies relevant concepts within CSK statements and systematically integrates them into ontology elements. This process enables the structured transformation of CSK into semantic rules aligned with ontologies, leveraging the owlready2[7] library for ontology manipulation. We use the Owlready2 library, a Python module for loading, editing, and reasoning with OWL ontologies, due to its seamless integration with Python-based systems and support for ontology-driven rule reasoning.

Formalizing Semantic Rule Specialization using Common Sense Knowledge (CSK)

Definition 1: Rule Template

A Rule Template, RT, is defined as a logical expression containing placeholders that represent general classes or relationships.

Example:

$$\text{process}(x) \rightarrow \exists y \text{process}(y) \wedge \text{comesAfter}(x, y) \tag{1}$$

Definition 2: Common Sense Knowledge (CSK)

A CSK is a natural language statement that provides specific knowledge about a particular case, such as classes or instances related to a process (e.g., painting). CSK is extracted from LLMs using the chain-of-thought prompt engineering method. The CSK serves as the source from which specific information is extracted to replace placeholders in the Rule Template.

Example 1: “The result of the painting process is a painted object.”
Example 2: “After the painting process, you do the drying process.”
Example 3: “The drying process involves a dryer machine.”

Definition 3: Concrete Rule

A Concrete Rule, CR, is derived from a Rule Template by replacing its placeholders with specific classes or instances extracted from a CSK.

Function: ***SpecializeRule***

Formally defined, the function SpecializeRule is:

$$\text{SpecializeRule} : RT \times CSK \rightarrow CR \tag{2}$$

Process:

1. **Input:** A Rule Template, RT, and Common Sense Knowledge, CSK.
2. **Extraction:** Identify and extract specific classes or instances from the CSK.
3. **Substitution:** Systematically replace the placeholders in RT with the extracted classes or instances.
4. **Output:** Return a Concrete Rule, CR.

[7] https://owlready2.readthedocs.io/en/v0.48/

Rule 1

Given the Rule Template RT based on standard vocabulary classes and property relations from BFO and IOF:

$$\text{IOF:MaterialProduct}(x) \rightarrow \exists y \text{BFO:process}(y) \wedge \text{IOF:isOutputOf}(x, y) \tag{3}$$

CSK: "The result of painting is a painted object."

Applying *SpecializeRule*:

$$\text{paintedObject}(x) \rightarrow \exists y \text{painting}(y) \wedge \text{IOF:isOutputOf}(x, y) \tag{4}$$

Rule 2

Given the Rule Template RT:

$$\text{BFO:process}(x) \rightarrow \exists y \text{BFO:process}(y) \wedge \text{BFO:precedes}(y, x) \tag{5}$$

CSK: "After painting process, you should perform a drying process."

Applying *SpecializeRule*:

$$\text{drying}(x) \rightarrow \exists y \text{painting}(y) \wedge \text{BFO:precedes}(y, x) \tag{6}$$

Rule 3

Given the Rule Template RT:

$$\text{BFO:process}(x) \rightarrow \exists y \text{MSDL:productionEquipment}(y) \wedge \text{BFO:participatesInAtSomeTime}(y, x) \tag{7}$$

CSK: "The drying process involves a dryer machine."

Applying *SpecializeRule*:

$$\text{drying}(x) \rightarrow \exists y \text{dryer}(y) \wedge \text{BFO:participatesInAtSomeTime}(y, x) \tag{8}$$

Explanation: The function *SpecializeRule* refines a general rule template based on CSK to yield a concrete rule suitable for a particular context. In this case, the rule template associates a process with the equipment involved in it. The CSK statement "The drying process involves a dryer machine" allows us to adapt the template to indicate that a dryer machine participates in the drying process.

Our method utilizes the pattern recognition capabilities of LLMs to interpret varied expressions of CSK and map them into predefined rule structures. LLMs are highly proficient at extracting and preserving the core semantics of patterns, even when expressed in diverse ways. This adaptability ensures that the system remains robust in handling the complex language often found in technical or industrial texts. The primary purpose of translating CSK into semantic rules is to ensure that ontology classes, subclasses, and predicates align with standard vocabularies. This alignment helps resolve LLMs' common issues with ontology-incompatible outputs. Our methodology employs `OWLReady2` for ontology manipulation. When CSK is converted into ontology elements through semantic rule generation, we verify whether the resulting classes and predicates match a standard vocabulary. If a close match is found based on definitions and axioms, the concept is used directly or mapped via a predefined semantic rule.

For instance, in the CSK statement, The result of painting is a painted object, the term "painting" aligns with the concept `BFO:Process`, while the painted object corresponds to `IOF:MaterialProduct`. Our semantic rule-based mapping establishes that a painted object is the output of a painting process using the IOF property `IOF:isOutputOf`.

5. Evaluation

To evaluate the effectiveness of our approach for transforming CSK into semantic rules aligned with standard ontologies, we conducted a two-part study focusing on (i) the correctness of class and relation extraction, (ii) semantic alignment with reference ontologies, and (iii) the practical usability of the generated rules for ontology population and reasoning.

We compiled a dataset of 50 CSK statements related to manufacturing processes (e.g., painting, drying, welding), extracted from LLM queries using chain-of-thought prompting. Each statement was processed using our `SpecializeRule` function to generate corresponding semantic rules. The evaluation was conducted in three stages:

1. **Class/Relation Extraction Accuracy**: Manually annotated gold-standard mappings of ontology classes and relations were created for the CSK statements.
2. **Semantic Alignment**: We assessed whether the generated rules used vocabulary terms consistent with BFO, IOF, and MSDL ontologies.
3. **Usability in Ontology Population**: We tested the generated rules in populating OWL ontologies via the `owlready2` API and evaluated their syntactic and semantic correctness.

5.1. Metrics

We used the following metrics:

- **Precision**: Fraction of correctly mapped classes/relations over all predicted.
- **Recall**: Fraction of correct mappings in the gold standard that were retrieved by the system.
- **Semantic Validity**: Percentage of rules whose predicates and classes corresponded to terms defined in BFO, IOF, or MSDL.
- **Rule Usability**: Proportion of rules successfully instantiated and executed within an OWL ontology environment.

Table 1
Evaluation Results over 50 CSK Statements

Metric	Value	Interpretation
Class Extraction Precision	0.86	Most predicted classes were correct
Class Extraction Recall	0.78	Some relevant classes missed
Relation Extraction Precision	0.90	High accuracy in relation identification
Relation Extraction Recall	0.82	Moderate coverage of all possible relations
Semantic Validity	92%	Rules used standard ontological terms
Rule Usability in OWL	88%	Rules instantiated without error

5.2. User Study: Expert Evaluation of Semantic Rules

To complement the quantitative metrics, we conducted a small-scale user study involving five domain experts from the fields of manufacturing and knowledge engineering. The participants were asked to evaluate a randomized subset of 20 generated semantic rules corresponding to CSK statements.

Each expert assessed the following dimensions on a 5-point Likert scale:

1. **Correctness**: Does the rule correctly represent the intended meaning of the CSK statement?
2. **Ontological Alignment**: Are the classes and predicates correctly aligned with standard vocabularies (BFO, IOF, MSDL)?
3. **Usefulness**: Would this rule be useful for automating ontology population or reasoning?

Results:

Table 2
Average Expert Ratings (Scale 1–5)

Evaluation Criterion	Mean Score	Std. Dev.
Correctness	4.4	0.48
Ontological Alignment	4.6	0.50
Usefulness	4.5	0.53

6. Application of Proposed Methodology in Manufacturing

We propose MACS-KG[8] [59], a specialized knowledge graph that incorporates Manufacturing Commonsense Knowledge (MCSK) to enhance reasoning and explainability within manufacturing decision-making processes. The core innovation of MACS-KG lies in its ability to extract domain-specific knowledge from pre-trained Large Language Models (LLMs) through Chain-of-Thought prompt engineering, leveraging established MCSK patterns [1] [59]. The extracted knowledge is then automatically transformed into First-Order Logic (FOL) representations that

[8]https://csk.chaikmat-anr.uttop.fr

align with standard ontological vocabularies such as the Basic Formal Ontology (BFO)[9], the Industrial Ontologies Foundry (IOF)[10], and the Machine Service Description Language (MSDL)[11]. This semantic alignment is critical for ensuring interoperability and consistent reasoning across heterogeneous manufacturing data sources. Once users validate the generated FOL rules, they are converted into executable SPARQL and Datalog rules.

The **MACS-KG user interface** provides two primary functionalities: Building MACS-KG: Users generate semantic rules grounded in MCSK using LLMs or manual rule templates (Fig. 1). After rule creation, the system allows users to save these rules in a graph database, ensuring structured storage and enabling efficient retrieval.

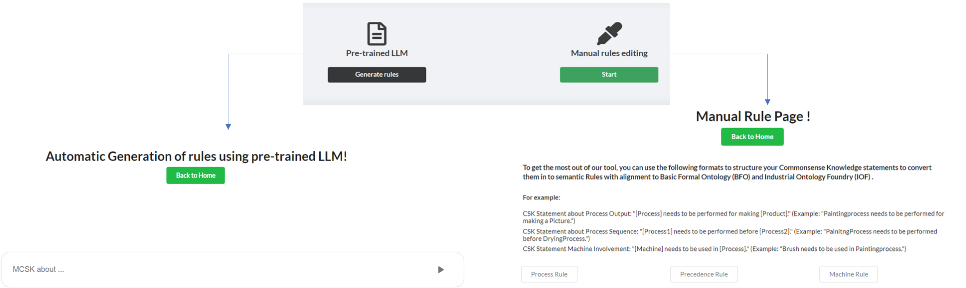

Figure 1: MACS-KG Building Using Semantic Rules

Exploring MACS-KG: Users can query the knowledge graph using SPARQL queries, visualize the graph structure, and manage stored rules through an interface intended to be intuitive and accessible for domain experts, though its usability would benefit from further evaluation. (Fig. 2).

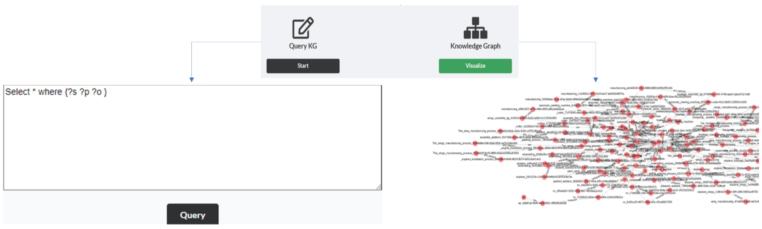

Figure 2: MACS-KG Exploration via SPARQL Querying

To demonstrate the practical utility of the MACS-KG framework, we applied it to a car manufacturing scenario (Fig.3). This use case illustrates how MACS-KG integrates knowledge of manufacturing processes and production equipment, allowing users to validate and refine generated semantic rules relevant to automotive production workflows. The system enforces an initial validation step where user-approved rules are required before they are committed to the knowledge graph. This validation prevents invalid or semantically inconsistent rules from

[9]https://basic-formal-ontology.org
[10]https://spec.industrialontologies.org/iof/
[11]https://labs.engineering.asu.edu/semantics/ontology-download/msdl-ontology/

entering the database and thereby protects the integrity of downstream SPARQL queries and reasoning tasks.

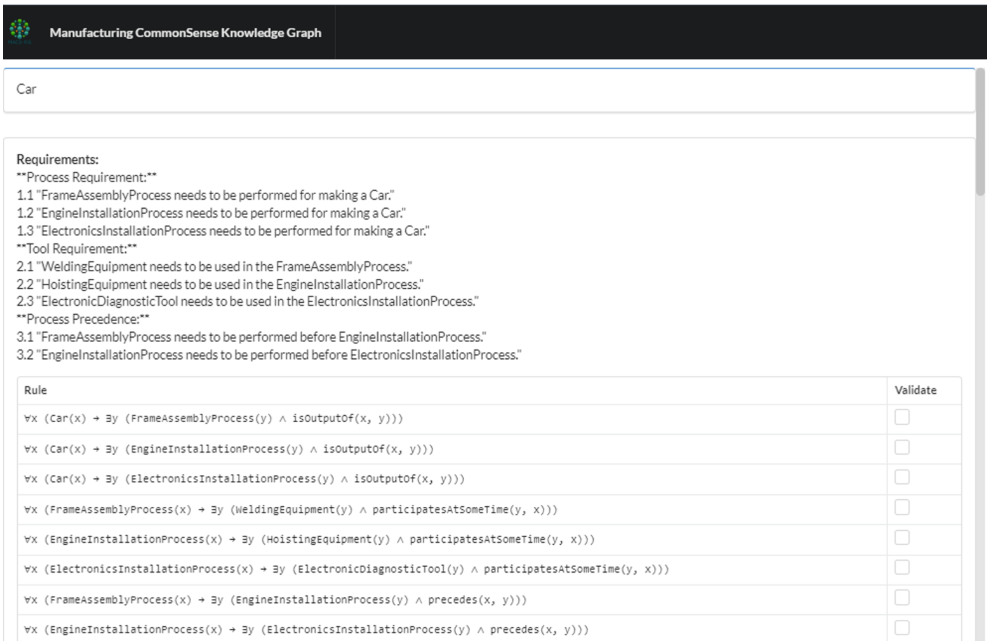

Figure 3: MACS-KG Use Case: Car Manufacturing

Figure 4 shows an example of validated rules encompassing car manufacturing processes and production equipment, as managed within the MACS-KG platform. This step ensures that only domain-relevant, ontologically aligned knowledge is incorporated, thereby enhancing the reliability of decision support derived from the knowledge graph.

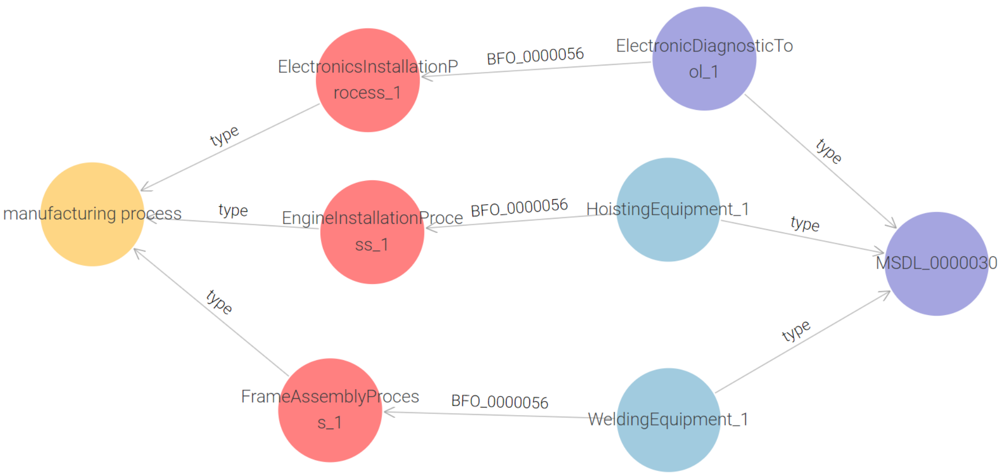

Figure 4: Validated Car Processes and Production Equipment in MACS-KG

7. Conclusion

Ontology development has traditionally required significant manual effort and domain expertise. This paper proposed a methodology for automating the transformation of commonsense knowledge (CSK), extracted via Large Language Models (LLMs), into semantically valid rules aligned with standard ontologies.

By bridging natural language processing and formal ontology engineering, our approach enables scalable, explainable, and ontology-aligned rule generation. Adhering to foundational vocabularies such as BFO, IOF, and MSDL, the method supports semantic interoperability and logical reasoning in OWL-based environments.

Future work will focus on expanding domain coverage, improving recall through prompt optimization, and incorporating human-in-the-loop strategies for validation. While demonstrated in the manufacturing domain, this approach lays the foundation for integrating LLM-based commonsense reasoning into broader semantic web applications.

Declaration on Generative AI

We acknowledge the use of the OpenAI API for generating experimental content and the GRaph DB and RDFox rule engine as the underlying graph database for reasoning tasks. Grammarly and ChatGPT were employed to assist in language refinement. However, the authors take full responsibility for the scientific content and research ideas presented in this study.

Disclaimer

Since September 2024, Mohamed Hedi Karray has joined European Innovation Council and SMEs Executive Agency. The views expressed in this publication are the responsibility of the authors and do not necessarily reflect the views of the European Commission nor of the European Innovation Council and SMEs Executive Agency. The European Commission or the European Innovation Council and SMEs Executive Agency are not liable for any consequence stemming from the reuse of this publication.

Acknowledgements:

This work is performed within the CHAIKMAT project funded by the French National Research Agency (ANR) under grant agreement " ANR-21-CE10-0004-01"

References

[1] Naqvi, S. M. R. (2025). Exploring LLMs and semantic XAI for industrial robot capabilities and manufacturing commonsense knowledge (Doctoral dissertation, Université de Toulouse).

[2] Schank, R.C. and Abelson, R.P., 2013. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Psychology Press.

[3] Carey, Susan, and Elizabeth Spelke. "Domain-specific knowledge and conceptual change." Mapping the mind: Domain specificity in cognition and culture 169 (1994): 200.
[4] Fensel, D. (2001). Ontologies. In Ontologies: A silver bullet for knowledge management and electronic commerce (pp. 11-18). Berlin, Heidelberg: Springer Berlin Heidelberg.
[5] Zang, L.J., Cao, C., Cao, Y.N., Wu, Y.M. and Cao, C.G., 2013. A survey of commonsense knowledge acquisition. Journal of Computer Science and Technology, 28(4), pp.689-719.
[6] Naqvi, M. R., Sarkar, A., Ameri, F., Araghi, S. N., & Karray, M. H. (2023). Application of MSDL in Modeling Capabilities of Robots. In CEUR Workshop Proceedings (Vol. 3595). CEUR-WS.
[7] Naqvi, Syed Muhammad Raza. Exploration des LLM et de l'XAI sémantique pour les capacités des robots industriels et les connaissances communes en matière de fabrication. Diss. Université de Toulouse (2023-....), 2025.
[8] Li, J., Hui, B., Qu, G., Yang, J., Li, B., Li, B., ... & Li, Y. (2023). Can llm already serve as a database interface? a big bench for large-scale database grounded text-to-sqls. Advances in Neural Information Processing Systems, 36, 42330-42357.
[9] Nguyen, T. P. (2024). Large-Scale Acquisition of Refined Commonsense Knowledge.
[10] Balakrishna, M. and Moldovan, D., 2013, May. Automatic building of semantically rich domain models from unstructured data. In The Twenty-Sixth International FLAIRS Conference.
[11] Tekli, J. (2016). An overview on xml semantic disambiguation from unstructured text to semi-structured data: Background, applications, and ongoing challenges. IEEE Transactions on Knowledge and Data Engineering, 28(6), 1383-1407.
[12] Graham, S., Yates, D. and El-Roby, A., 2023. Investigating antiquities trafficking with generative pre-trained transformer (GPT)-3 enabled knowledge graphs: A case study. Open Research Europe, 3, p.100.
[13] Yenduri, G., Ramalingam, M., Chemmalar Selvi, G., Supriya, Y., Srivastava, G., Maddikunta, P.K.R., Deepti Raj, G., Jhaveri, R.H., Prabadevi, B., Wang, W. and Athanasios, V., GPT (Generative Pre-trained Transformer)–A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions.
[14] Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y. and Gašević, D., 2024. Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(**1**), pp.90-112.
[15] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E. and Krusche, S., 2023. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, **103**, p.102274.
[16] Zhou, J.P., Staats, C.E., Li, W., Szegedy, C., Weinberger, K.Q. and Wu, Y., 2023, October. Don't Trust: Verify–Grounding LLM Quantitative Reasoning with Autoformalization. In The Twelfth International Conference on Learning Representations.
[17] Xie, C., Chen, C., Jia, F., Ye, Z., Shu, K., Bibi, A., Hu, Z., Torr, P., Ghanem, B. and Li, G., 2024. Can Large Language Model Agents Simulate Human Trust Behaviors?. arXiv preprint arXiv:2402.04559.
[18] Choudhury, M. and Deshpande, A., 2021, May. How Linguistically Fair Are Multilingual Pre-Trained Language Models?. In Proceedings of the AAAI conference on artificial intelligence

(Vol. 35, No. **14**, pp. 12710-12718).
[19] Zhang, J., Bao, K., Zhang, Y., Wang, W., Feng, F., & He, X. (2023, September). Is chatgpt fair for recommendation? evaluating fairness in large language model recommendation. In Proceedings of the 17th ACM Conference on Recommender Systems (pp. 993-999).
[20] Neuhaus, F., 2023. Ontologies in the era of large language models–a perspective. Applied Ontology, 18(**4**), pp.399-407.
[21] Kripke SA. The question of logic. Mind. 2024 Jan 1;**133**(529):1-36.
[22] Zangwill N. Does knowledge depend on truth? Acta Analytica. 2013 Jun;**28**(2):139-44.
[23] Dupuy JP. Common knowledge, common sense. Theory and Decision. 1989 Jul;**27**(1-2):37-62.
[24] Berg-Cross G. Commonsense and Explanation. Journal of the Washington Academy of Sciences. 2020 Dec 1;**106**(4):39-66.
[25] Lenat DB. From 2001 to 2001: Common Sense and the Mind of HAL. HAL's Legacy. 2001:193-209.
[26] Dhamani N, Engler M. Introduction to Generative AI. Simon and Schuster; 2024 Feb 27.
[27] Han, Mengjie, et al. "Perspectives of Machine Learning and Natural Language Processing on Characterizing Positive Energy Districts." Buildings **14.2** (2024): 371.
[28] Barbierato, Enrico, and Alice Gatti. "The Challenges of Machine Learning: A Critical Review." Electronics **13**, no. 2 (2024): 416.
[29] Retzlaff, Carl Orge, Srijita Das, Christabel Wayllace, Payam Mousavi, Mohammad Afshari, Tianpei Yang, Anna Saranti, Alessa Angerschmid, Matthew E. Taylor, and Andreas Holzinger. "Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities." Journal of Artificial Intelligence Research 79 (2024): 359-415.
[30] Long, S., Tan, J., Mao, B., Tang, F., Li, Y., Zhao, M., & Kato, N. (2025). A survey on intelligent network operations and performance optimization based on large language models. IEEE Communications Surveys & Tutorials.
[31] Kirk, H. R., Vidgen, B., Röttger, P., & Hale, S. A. (2024). The benefits, risks and bounds of personalizing the alignment of large language models to individuals. Nature Machine Intelligence, 6(4), 383-392.
[32] Korinek A. Language models and cognitive automation for economic research. National Bureau of Economic Research; 2023 Feb 13.
[33] Babaei Giglou H, D'Souza J, Auer S. LLMs4OL: Large language models for ontology learning. In International Semantic Web Conference 2023 Oct 27 (pp. 408-427). Cham: Springer Nature Switzerland.
[34] Pan JZ, Razniewski S, Kalo JC, Singhania S, Chen J, Dietze S, Jabeen H, Omeliyanenko J, Zhang W, Lissandrini M, Biswas R. Large language models and knowledge graphs: Opportunities and challenges. arXiv preprint arXiv:2308.06374. 2023 Aug 11.
[35] Koubaa A, Boulila W, Ghouti L, Alzahem A, Latif S. Exploring ChatGPT Capabilities and Limitations: A Survey. IEEE Access. 2023 Oct 23.
[36] Babaei Giglou, H., D'Souza, J., & Auer, S. (2023, October). LLMs4OL: Large language models for ontology learning. In International Semantic Web Conference (pp. 408-427). Cham: Springer Nature Switzerland.
[37] Burtsev M, Reeves M, Job A. The Working Limitations of Large Language Models. MIT Sloan Management Review. 2024;**65**(2):8-10.

[38] Molinari, A., & Sandri, S. (2024, November). Evolution of lms design and implementation in the age of ai and large language models. In Proceedings of the Second International Workshop on Artificial INtelligent Systems in Education co-located with 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024), Bolzano, Italy.

[39] Li Y, Huang Y, Lin Y, Wu S, Wan Y, Sun L. I Think, Therefore I am: Awareness in Large Language Models. arXiv preprint arXiv:2401.17882. 2024 Jan 31. Zangwill N. Does knowledge depend on truth?. Acta Analytica. 2013 Jun;**28**(2):139-44.

[40] Jansen L. Categories: The top-level ontology. Applied ontology: An introduction. 2008 Jan:173-96.

[41] Schmidt, Douglas C., Jesse Spencer-Smith, Quchen Fu, and Jules White. "Towards a catalog of prompt patterns to enhance the discipline of prompt engineering." ACM SIGAda Ada Letters 43, no. 2 (2024): 43-51.

[42] Li, J., Tang, T., Zhao, W. X., Nie, J. Y., & Wen, J. R. (2024). Pre-trained language models for text generation: A survey. ACM Computing Surveys, 56(9), 1-39.

[43] Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., Agirre, E., Heintz, I. and Roth, D., 2023. Recent advances in natural language processing via large pre-trained language models: A survey. ACM Computing Surveys, 56(2), pp.1-40.

[44] Ma, K., Grandi, D., McComb, C., & Goucher-Lambert, K. (2023, August). Conceptual design generation using large language models. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 87349, p. V006T06A021). American Society of Mechanical Engineers.

[45] Pan S, Luo L, Wang Y, Chen C, Wang J, Wu X. Unifying large language models and knowledge graphs: A roadmap. IEEE Transactions on Knowledge and Data Engineering. 2024 Jan 10.

[46] Shafee, S., Bessani, A., & Ferreira, P. M. (2025). Evaluation of LLM-based chatbots for OSINT-based Cyber Threat Awareness. Expert Systems with Applications, 261, 125509.

[47] Acharya, K., Velasquez, A., & Song, H. H. (2024). A survey on symbolic knowledge distillation of large language models. IEEE Transactions on Artificial Intelligence.

[48] 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, 36(7), 3580-3599.

[49] Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2025). The rise and potential of large language model based agents: A survey. Science China Information Sciences, 68(2), 121101.

[50] Saeedizade, M. J., & Blomqvist, E. (2024, May). Navigating ontology development with large language models. In European Semantic Web Conference (pp. 143-161). Cham: Springer Nature Switzerland.

[51] Andročec, D. (2025). Using Large Language Models for Ontology Development. Engineering Proceedings, 104(1), 9.

[52] Joachimiak, M. P., Miller, M. A., Caufield, J. H., Ly, R., Harris, N. L., Tritt, A., ... & Bouchard, K. E. (2024). The Artificial Intelligence Ontology: LLM-assisted construction of AI concept hierarchies. Applied Ontology, 19(4), 408-418.

[53] García-Fernández, J., Verhoosel, J., Ubacht, J., & Bakker, R. M. (2025). Ontology Engineering with Large Language Models: Unveiling the potential of human-LLM collaboration in the

ontology extension process. extraction, 7, 15.
[54] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A. and Fung, P., 2023. Survey of hallucination in natural language generation. ACM Computing Surveys, **55**(12), pp.1-38.
[55] Zhang, H., Song, H., Li, S., Zhou, M. and Song, D., 2023. A survey of controllable text generation using transformer-based pre-trained language models. ACM Computing Surveys, **56**(3), pp.1-37.
[56] Li, J., Garijo, D., & Poveda-Villalón, M. (2025). Large Language Models for Ontology Engineering: A Systematic Literature Review.
[57] Mai, Huu Tan, Cuong Xuan Chu, and Heiko Paulheim. "Do LLMs really adapt to domains? an ontology learning perspective." International Semantic Web Conference. Cham: Springer Nature Switzerland, 2024.
[58] Naqvi, M. R., Sarkar, A., Ameri, F., Elmhadhbi, L., & Karray, M. H. (2025, June). MACS-KG: MAnufacturing CommonSense Knowledge Graph. In European Semantic Web Conference (pp. 120-124). Cham: Springer Nature Switzerland.
[59] Naqvi, M. R., Sarkar, A., Ameri, F., Elmhadhbi, L., & Karray, M. H. (2024, December). Manufacturing Commonsense Knowledge. In International Knowledge Graph and Semantic Web Conference (pp. 320-333). Cham: Springer Nature Switzerland.