A Web Engineering Method for AI-Assisted Knowledge Graph Construction in Industrial Domains

Maheshika Hansamalee Walpola[1 [0009-0002-9362-0278]], Sheeba Samuel[1 [0000-0002-7981-8504]], and Martin Gaedke[1 [0000-0002-6729-2912]]

Chemnitz University of Technology, Chemnitz, Germany
{maheshika.walpola,sheeba.samuel,martin.gaedke}@informatik.tu-chemnitz.de

Abstract. Predictive maintenance systems rely on machine learning (ML) to anticipate equipment failures, but domain engineers often cannot explain why a prediction was made, limiting trust and adoption. Knowledge graphs (KGs) can connect predictions with structured domain knowledge, but building industrial KGs requires semantic web expertise domain engineers lack. This PhD proposal addresses three problems: ML predictions lack human-readable explanations, KG construction has no reproducible method with formal validation, and no web-based approach enables non-technical users to build and query KGs. The proposed Web Engineering method integrates large language model (LLM) assistance across the KG lifecycle, from ontology elicitation and data validation to natural language querying and explanation generation. The method is realised through a web portal following End-User Development (EUD) principles, enabling domain engineers to construct ontologies, ingest data, query the KG, and inspect prediction explanations. Evaluation uses CMP semiconductor manufacturing, with a second use case planned in wind turbine monitoring.

Keywords: Knowledge Graphs · Web Engineering · AI Prediction Explanation · Ontology Engineering · Predictive Maintenance · End-User Development · Industry 5.0

1 Introduction

Predictive maintenance systems powered by machine learning (ML) can detect early signs of equipment failure from sensor data [1], and knowledge graphs (KGs) can make those predictions traceable by linking them to structured domain knowledge [2,3]. Building such systems for industrial use, however, remains an open challenge [5].

Industry 5.0 calls for human-centred manufacturing where domain engineers, not only IT specialists, take part in shaping AI-driven processes [13]. At present, no method enables these engineers to construct and use knowledge structures independently. This PhD proposal presents a Web Engineering method that uses AI-assisted ontology construction, SHACL (Shapes Constraint Language, a W3C standard for validating RDF data)-based validation, and a web portal that enables domain engineers to construct, query, and obtain explanations from industrial KGs through a standard browser.

1.1 Problem Statement and Motivation

Three problems limit the adoption of KG-based explanation systems in industrial predictive maintenance.

PROB 1: ML predictions are not interpretable for domain engineers. When a predictive maintenance system warns that a machine is likely to fail, the engineer needs to understand *why* before acting [1]. Existing KG approaches can link predictions to process context [2], but they are built and operated by knowledge engineers, not by the domain experts who actually need them.

PROB 2: Industrial KG construction lacks reproducible methods. Building industrial KGs has typically taken months of specialised effort, and neither reproducible construction methods nor formal sensor data validation have been established [5]. Although large language models (LLMs) can now assist across the KG construction lifecycle [12], existing LLM-assisted approaches [6,7] have not been applied in industrial sensor domains with formal quality checks.

PROB 3: No web-based method exists for non-technical users. Current KG tools such as Protégé require knowledge of OWL, SPARQL, and RDF, which manufacturing and process engineers typically do not have [10]. These engineers possess deep domain knowledge of equipment and failure modes but lack the semantic web skills to author ontologies or formulate graph queries. To date, no Web Engineering method offers a documented workflow that enables such users to construct industrial KGs, enforce data quality through SHACL validation, and obtain explanations for ML predictions through a single process with clearly assigned roles, defined inputs and outputs, and evaluation criteria at each step.

1.2 Research Aims and Objectives

This research is guided by the following overarching question: *How should a Web Engineering method be designed to support AI-assisted construction, quality assurance, and explanation of industrial knowledge graphs within a single reproducible process accessible to domain engineers?*

The research addresses this gap through four objectives. **OBJ 1:** To develop and evaluate an LLM-assisted pipeline for industrial OWL 2 ontology construction, benchmarked against manual construction in coverage, correctness, and time. **OBJ 2:** To construct and validate RDF-based industrial KGs through automated sensor data mapping with SHACL constraints, measuring conformance rates on real sensor datasets. **OBJ 3:** To design and evaluate a web portal following EUD principles that enables domain engineers to perform all method stages and inspect prediction explanations through a browser. **OBJ 4:** To assess cross-domain generalisability by replicating the method in a structurally different industrial domain. The research is conducted within the WiProFlex industry project[1], with CMP semiconductor manufacturing as the primary evaluation

[1] https://www.tu-chemnitz.de/wiproflex/index.html.en

domain. A second case study in wind turbine condition monitoring will test generalisability; wind turbines differ from CMP in sensor types, failure modes, and maintenance workflows, and an existing KG baseline (XAI4Wind [4]) allows direct comparison.

2 Related Work

Research relevant to this PhD spans four areas. Each provides foundations but leaves the central problem unsolved.

Structured domain knowledge can ground ML outputs in human-understandable concepts [2], and a survey establishes that KGs can serve as explanation tools across ML systems [3]. In industrial maintenance, the XAI4Wind system demonstrated a multimodal KG for explainable wind turbine decision support [4], and a review examined KG applications across production environments [5]. These works demonstrate the value of KG-based explanations but all required extensive manual effort by knowledge engineers, none presented a reusable construction method, and none evaluated explanation quality with domain experts.

A recent roadmap shows that LLMs can assist across the full KG construction lifecycle, including ontology engineering, knowledge extraction, graph completion, and question answering [12]. More specifically, LLM-assisted pipelines have produced ontologies with accuracy approaching that of human experts [6], and LLMs can perform ontology learning tasks such as term typing and taxonomy discovery [7]. These efforts focus on general knowledge engineering and have not been tested with SHACL-enforced validation or human-in-the-loop review in industrial sensor domains. A SHACL-based web-form generator has been shown to reduce editing errors for non-specialist users [8], though it does not cover ontology construction, LLM assistance, or ML prediction explanation.

End-User Development (EUD) research investigates how people without programming skills can build and extend software systems [9]. Recent work on accelerating ontology engineering with LLMs indicates that these tasks still require significant expert labour and calls for approaches that lower the barrier for non-specialists [10], which motivates the portal design proposed here.

To the best of our knowledge, this is the first method to bring together LLM-assisted ontology construction, SHACL-based validation, a web-based EUD portal, and evaluation with domain engineers in a single reproducible Web Engineering workflow.

3 Research Methodology

This section presents the four stages of the method and the planned evaluation.

3.1 Overall Approach

The research follows Design Science Research (DSR) [11] and produces a reusable method organised as a structured engineering lifecycle of requirements, design, implementation, and evaluation. The method comprises four stages, visualised in Fig. 1 and linked to the objectives from Section 1.2.

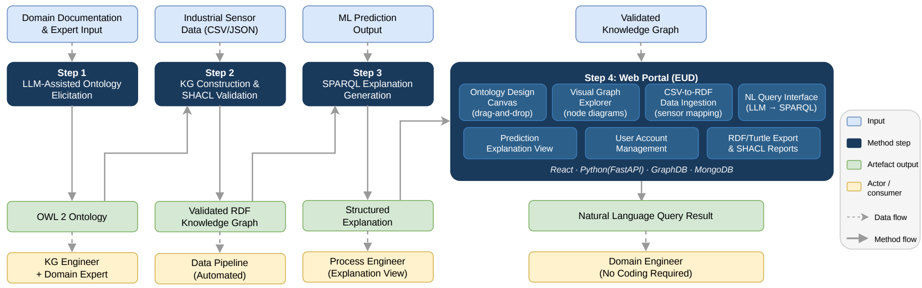

Fig. 1. Overview of the four-stage method.

As shown in Fig. 1, in the first stage (OBJ 1), the knowledge engineer and domain expert extract domain knowledge with LLM assistance. Documentation such as sensor manuals, equipment specifications, and maintenance logs is processed by the LLM using structured prompts that guide the extraction of OWL 2 class hierarchies, properties, and relationships. Every element the LLM proposes is reviewed by the domain expert before it is accepted into the ontology.

In the second stage (OBJ 2), the knowledge graph is constructed and validated. First, the knowledge engineer and domain expert define SHACL constraint sets that match the ontology from Stage 1. An automated pipeline then maps industrial sensor data (CSV or JSON) to the OWL 2 classes defined in Stage 1 and stores the readings as RDF triples with provenance metadata, producing the knowledge graph. Before the graph is used, SHACL validation checks for property violations, datatype mismatches, and missing values.

The third stage (OBJ 3) uses the validated knowledge graph to generate explanations for ML predictions. When the ML model flags a prediction, the system selects a parameterised SPARQL query template matching the prediction type (e.g., tool wear, slurry degradation) and retrieves a subgraph of related process steps, sensor readings, and failure patterns. The LLM then summarises this retrieved subgraph into a domain-grounded, plain-language explanation for the domain engineer, ensuring that the output is constrained to factual KG content rather than free generation.

The fourth stage (OBJ 1-4) brings all previous stages together in a web portal built on React, FastAPI, and GraphDB (see Fig. 1, Step 4). Through a standard browser, practitioners can edit ontologies using drag and drop, ingest sensor data via CSV-to-RDF upload with automated column mapping, query the graph using natural language through LLM-to-SPARQL translation, inspect prediction explanations, and explore the graph visually. The design separates schema authoring from instance data entry and shows validation feedback inline.

3.2 Evaluation Design

The primary evaluation uses CMP manufacturing data from the WiProFlex project, with real sensor data and direct access to process engineers. For OBJ 1, a controlled experiment compares AI-assisted ontology construction against manual construction in Protégé, measuring time, coverage, and correctness. For

OBJ 2, SHACL conformance rates are measured on the CMP sensor dataset. For OBJ 3, a usability study with process engineers will evaluate the portal in terms of usability, trust, and overall user experience, combining qualitative and quantitative measures as appropriate. For OBJ 4, the method is replicated in the wind turbine domain and compared against XAI4Wind on construction time and explanation quality.

4 Preliminary Results

The core method components are under active development and testing in the CMP domain.

CMP Process Ontology (OBJ 1). A CMP Process Ontology (CMPO) is being constructed with domain experts and currently comprises 57 OWL 2 classes, 32 object properties, and 33 data properties covering process stages, sensor types, polishing parameters, slurry chemistry, wafer properties, and failure modes. A systematic literature review established that no dedicated CMP process ontology existed prior to this work.

RDF Pipeline and SHACL Validation (OBJ 2). A data pipeline currently maps 28 CMP sensor columns to CMPO classes and stores readings as RDF triples with provenance metadata. An initial SHACL validation run found violations in three property categories; these were resolved through schema alignment with the domain expert, bringing conformance to an acceptable level. Further constraint coverage is being extended as part of Phase 2.

KG Web Portal (OBJ 3). An early version of the portal supports the core method stages through a standard browser, including ontology editing, data ingestion, natural language querying, prediction explanation, and graph exploration.[2]

5 Research Plan and Contributions

Research Plan. Phase 1 (in progress, OBJ 1 and OBJ 3): literature review, requirements elicitation with CMP domain experts, initial CMPO construction, and portal development. Phase 2 (mid-2026, OBJ 2): extension of SHACL coverage to the full CMP sensor schema and completion of the explanation module. Phase 3 (Q3–Q4 2026, OBJ 1 and OBJ 3): formal usability study with process engineers and controlled ontology construction experiment. Phase 4 (2027, OBJ 4): method replication in wind turbine monitoring.

Contributions to Web Engineering. This research contributes a structured method for building industrial AI systems that provide human-readable prediction explanations, through web-based KG engineering and end-user interaction. It will produce: **(a)** a reusable Web Engineering method for AI-assisted industrial KG construction with documented stages, roles, and artefacts; **(b)** a validated evaluation framework combining controlled experiments, SHACL-based quality metrics, and usability studies; **(c)** an EUD web portal enabling practitioners

[2] Portal prototype: https://github.com/MaheshikaWalpola/KGPortal

to construct, validate, query, and obtain explanations from industrial KGs; and **(d)** cross-domain empirical evidence from CMP and wind turbine domains showing that the method generalises. Method artefacts, prompt templates, and SHACL patterns are available in the project repository.

The method aims to make industrial KG construction accessible through the web, bridging ML-driven predictions and the human understanding needed to act on them. Results will be disseminated at Web Engineering and Semantic Web venues.

References

1. Carvalho, T.P., et al.: A systematic literature review of machine learning methods applied to predictive maintenance. Comput. Ind. Eng. **137**, 106024 (2019). https://doi.org/10.1016/j.cie.2019.106024
2. Lécué, F.: On the Role of Knowledge Graphs in Explainable AI. Semant. Web **11**(1), 41–51 (2020). https://doi.org/10.3233/SW-190374
3. Tiddi, I., Schlobach, S.: Knowledge Graphs as Tools for Explainable Machine Learning: A Survey. Artif. Intell. **302**, 103627 (2022). https://doi.org/10.1016/j.artint.2021.103627
4. Chatterjee, J., Dethlefs, N.: XAI4Wind: A Multimodal Knowledge Graph Database for Explainable Decision Support in Operations & Maintenance of Wind Turbines. arXiv:2012.10489 (2020). https://doi.org/10.48550/arXiv.2012.10489
5. Buchgeher, G., et al.: Knowledge Graphs in Manufacturing and Production: A Systematic Literature Review. IEEE Access **9**, 55537–55554 (2021). https://doi.org/10.1109/ACCESS.2021.3070395
6. Kommineni, V.K., König-Ries, B., Samuel, S.: From Human Experts to Machines: An LLM Supported Approach to Ontology and Knowledge Graph Construction. arXiv:2403.08345 (2024). https://doi.org/10.48550/arXiv.2403.08345
7. Giglou, H.B., D'Souza, J., Auer, S.: LLMs4OL: Large Language Models for Ontology Learning. In: Payne, T.R., et al. (eds.) ISWC 2023, LNCS, vol. 14265, pp. 408–427. Springer (2023). https://doi.org/10.1007/978-3-031-47240-4_22
8. Wright, J., et al.: Schímatos: A SHACL-Based Web-Form Generator for Knowledge Graph Editing. In: Pan, J.Z., et al. (eds.) ISWC 2020, LNCS, vol. 12507, pp. 65–80. Springer (2020). https://doi.org/10.1007/978-3-030-62466-8_5
9. Fischer, G.: End-User Development: Empowering Stakeholders with Artificial Intelligence, Meta-Design, and Cultures of Participation. In: Fogli, D., et al. (eds.) IS-EUD 2021, LNCS, vol. 12724, pp. 1–16. Springer (2021). https://doi.org/10.1007/978-3-030-79840-6_1
10. Shimizu, C., et al.: Accelerating Knowledge Graph and Ontology Engineering with Large Language Models. Web Semant. **84**, 100848 (2025). https://doi.org/10.1016/j.websem.2025.100848
11. Hevner, A.R., et al.: Design Science in Information Systems Research. MIS Q. **28**(1), 75–105 (2004). https://doi.org/10.2307/25148625
12. Pan, S., et al.: Unifying Large Language Models and Knowledge Graphs: A Roadmap. IEEE Trans. Knowl. Data Eng. **36**(7), 3580–3599 (2024). https://doi.org/10.1109/TKDE.2024.3352100
13. Breque, M., De Nul, L., Petridis, A.: Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry. European Commission (2021). https://doi.org/10.2777/308407