



NARRATE

Regenerative Resilient Smart Manufacturing Networks

D4.2(A) END-TO-END AI-DRIVEN VISIBILITY MODEL & SUPPORT DSS (A)

2025/03/10



Grant Agreement No.: 101138094
 Topic: HORIZON-CL4-2023-TWIN-TRANSITION-01-07
 Call: HORIZON-CL4-2023-TWIN-TRANSITION-01
 Type of action: HORIZON-RIA

D4.2(a) END-TO-END AI-DRIVEN VISIBILITY MODEL & SUPPORT DSS (a)

| | |
|------------------|---|
| Work package | WP4 O4.3, O4.4, O4.6 |
| Task | Task Number |
| Due date | 28/02/2025 |
| Submission date | 10/03/2025 |
| Deliverable lead | NUNSYS SA |
| Version | 1.0 |
| Authors | González, Tomás; Correa, David (UNN) |
| Reviewers | 1st reviewer - INSA, 2nd reviewer (BUL) |
| Abstract | The NARRATE project seeks to improve supply chain visibility and decision-making by implementing an AI-driven visibility model within Smart Manufacturing Networks. Central to the project is the Intelligent Manufacturing Custodian, which orchestrates supply chain processes through AI-based automation, ensuring optimal workflow. Key innovations include the Neuro-Symbolic Question Answering module, which enhances reasoning and transparency through graph-based ontologies, and a Rules Module that ensures consistent decision-making and compliance. The API Module facilitates seamless data exchange and system integration, enabling interoperability among various systems. NARRATE aims to strengthen supply chain resilience, optimize resource utilization, and create a more adaptive and intelligent manufacturing ecosystem. |
| Keywords | AI-driven visibility model; Intelligent Manufacturing Custodian (IMC); Supply chain resilience; Neuro-Symbolic Question Answering (NSQA); API Module |

Project co-funded by the European Commission in the H2020 Programme

| | |
|----------------------------|----------------------|
| Nature of the deliverable: | R - Document, report |
|----------------------------|----------------------|

| | |
|----------------------------|-------------|
| Nature of the deliverable: | PU - Public |
|----------------------------|-------------|

DOCUMENT REVISION HISTORY

| Version | Date | Description of change | List of contributor(s) |
|---------|------------|-----------------------------|------------------------|
| 0.1 | 2025/02/15 | First version | TG, DC |
| 0.2 | 2025/03/02 | Peer review | |
| 1.0 | 2025/03/10 | Final Review by coordinator | |

STATEMENT ON MAINSTREAMING GENDER

The NARRATE consortium is committed to including gender and intersectionality as a transversal aspect in the project’s activities. In line with EU guidelines and objectives, all partners – including the authors of this deliverable – recognise the importance of advancing gender analysis and sex-disaggregated data collection in the development of scientific research. Therefore, we commit to paying particular attention to including, monitoring, and periodically evaluating the participation of different genders in all activities developed within the project, including workshops, webinars and events but also surveys, interviews and research, in general. While applying a non-binary approach to data collection and promoting the participation of all genders in the activities, the partners will periodically reflect and inform about the limitations of their approach. Through an iterative learning process, they commit to plan and implement strategies that maximise the inclusion of more intersectional perspectives in their activities.

DISCLAIMER

Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

COPYRIGHT NOTICE

© NARRATE Consortium, 2023 - All rights reserved

No part of this publication may be translated, reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the publisher or provided the source is acknowledged.



Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

How to cite this report: NARRATE(2024). D4.2(a) End-to-end AI-driven visibility model & support DSS (a).

The NARRATE Consortium is the following:

| Participant number | Participant organization name | Short name | Country |
|--------------------|--|------------|---------|
| 1 | INSTITUTO TECNOLOGICO METALMECANICO, MUEBLE, MADERA, EMBALAJE Y AFINES-AIDIMME | AID | ES |
| 2 | SCIENTIFIC ACADEMY FOR SERVICE TECHNOLOGY EV | SERV | DE |
| 3 | FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV | FhG | DE |
| 4 | INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON | INSA | FR |
| 4.1 | INSAVALOR SA | INSA-V | FR |
| 5 | SOFTWARE AG | SAG | DE |
| 6 | F6S NETWORK IRELAND LIMITED | F6S | IE |
| 7 | SYNESIS-SOCIETA CONSORTILE A RESPONSABILITA LIMITATA | SYN | IT |
| 8 | MEDITERRANEAN WOOD FACTORY S.L. - MEDWOOD | MED | ES |
| 9 | DHL EXEL SUPPLY CHAIN SPAIN SL | DHL | ES |
| 10 | NUNSYS SA | NUN | ES |
| 12 | BUDATEC GMBH | BUD | DE |
| 12 | BRUNEL UNIVERSITY LONDON | BUL | UK |
| 13 | POLICY LAB OU | PL | EE |

Table of Contents

- LIST OF FIGURES 5
- ABBREVIATIONS 6
- 1. INTRODUCTION 8
 - 1.1. TASKS, DELIVERABLES AND MILESTONES 8
- 2. ARCHITECTURE 9
 - 2.1. COMPONENTS AND RELATIONSHIPS 9
 - 2.2. INTELLIGENT MANUFACTURING CUSTODIAN (IMC) 10
 - 2.2.1. AGENT-BASED ORCHESTRATION WITH LARGE LANGUAGE MODELS 10
 - 2.3. NEURO-SYMBOLIC QUESTION ANSWERING (NSQA) 12
 - 2.3.1. ONTOLOGY WEB LANGUAGE (OWL) 13
 - 2.3.2. NEO4J GRAPH DATABASE 14
 - 2.4. API CONNECTION MODULE 15
 - 2.4.1. KNOWLEDGE MODELS IN A HYBRID FEDERATED DATA MESH (D3.1) 15
 - 2.4.2. DIGITAL TWIN REPRESENTATION IN MANUFACTURING NETWORKS (D3.2) 16
 - 2.4.3. INTEGRATION OF LANGCHAIN AND OPENAPI STANDARDS 17
 - 2.4.4. API-DRIVEN INTEGRATION OF TOOLS WITHIN THE IMC 19
 - 2.4.5. TOOLS AND PROPOSED API INTEGRATION 20
 - 2.4.6. API INTEROPERABILITY FRAMEWORK 22
 - 2.5. RULES MODULE 22
 - 2.5.1. THE ROLE OF THE RULES MODULE IN LLM-BASED ORCHESTRATION 22
 - 2.5.2. ENHANCING AGENT-BASED ORCHESTRATION WITH STRUCTURED PROMPTS 22
- 3. THEORETICAL PROOF OF CONCEPT 24
 - 3.1. DESCRIPTION OF MODULES CONFIGURATIONS 24
 - 3.2. CONCEPTUAL WORKFLOW: SUPPLIER QUERY RESPONSE PROCESS 25
- 4. CONCLUSIONS 28
- 5. APPENDIX 29
 - Appendix 1: Large Language Models 29
 - Appendix 2: Graph ontology extended example 30
- 6. REFERENCES 31

LIST OF FIGURES

Figure 1: Architecture diagram 9

Figure 2: LLM agent orchestration..... 11

Figure 3: Graph ontology for supply chain example 12

Figure 4: OpenAPI v3.0 structure. source: <https://www.apideck.com/blog/introduction-to-openapi-specification> 18

Figure 5: Example of IMC using an API to retrieve weather data..... 19

Figure 6: Example of graph database used for the proof of concept..... 24

Figure 7: Proof of concept sequence diagram 27

Figure 8: Example of ontology..... 30

ABBREVIATIONS

| | |
|-------|-------------------------------------|
| API | Application Programming Interface |
| DSS | Decision Support System |
| DRL | Deep Reinforcement Learning |
| DT | Digital Twin |
| DTPL | Digital Twin Processing Language |
| EC | European Commission |
| ERP | Enterprise resource planning |
| FAISS | Facebook AI Similarity Search |
| GNN | Graph Neural Networks |
| IMC | Intelligent Manufacturing Custodian |
| IoT | Internet of things |
| JSON | JavaScript Object Notation |
| LLM | Large Language Models |
| MADQN | multi-agent deep Q-networks |
| MAS | multi-agent systems |
| NSQA | Neuro-Symbolic Question Answering |
| OAS | OpenAPI Specification |
| OWL | Ontology Web Language |
| PLM | Product lifecycle management |
| RDF | Resource Description Framework |
| RL | Reinforcement Learning |
| SOTA | State of the art |
| SMN | Smart Manufacturing Network |
| Tx.x | Task x.x |
| WP | Work Package |
| XML | extensible markup language |
| YAML | yet another markup language |

EXECUTIVE SUMMARY

The NARRATE project aims to enhance supply chain visibility and decision-making through an AI-driven visibility model. This model integrates key modules to improve efficiency and automation within Smart Manufacturing Networks (SMNs).

The Intelligent Manufacturing Custodian (IMC) is responsible for orchestrating supply chain processes by leveraging AI-based decision-making and automation. It acts as the central coordinator, ensuring smooth operations and optimal workflow execution across the supply chain.

The Neuro-Symbolic Question Answering (NSQA) module enhances reasoning capabilities through structured knowledge representation and graph-based ontologies. It provides context-aware insights by analysing complex relationships between supply chain components, improving transparency and informed decision-making.

The Rules Module defines operational constraints, ensuring consistent decision-making and regulatory compliance. It establishes guidelines that govern automated processes, preventing inefficiencies and aligning actions with predefined business objectives.

Lastly, the API Module facilitates seamless data exchange, enabling integration across various systems. It serves as the interface between different components, allowing real-time data flow and interoperability among digital tools, manufacturing systems, and external stakeholders.

Key innovations include:

- AI-driven orchestration to enhance supply chain efficiency.
- Graph-based ontologies to improve data interoperability and contextual awareness.
- A structured Rules Module for consistent decision-making.
- An API Module ensuring seamless data exchange and system integration.

By implementing these advancements, the project strengthens supply chain resilience, optimizes resource utilization, and fosters a more intelligent and adaptive manufacturing ecosystem.

1. INTRODUCTION

The development of knowledge models for Smart Manufacturing Network (SMNs) plays a crucial role in enhancing transparency, efficiency, and decision-making across the entire supply chain ecosystem. This project, specifically within the framework of D4.2(a) and T4.2, aims to extend the neuro-symbolic Decision Support System (DSS) outlined in D3.3 to create comprehensive knowledge models. These models encapsulate the intricate relationships and structures of SMNs, incorporating suppliers, manufacturers, logistics, providers, and customers. By using digital twin (DT) models and production process orchestration, this initiative seeks to provide an in-depth understanding of supply chain operations and facilitate optimization.

1.1. TASKS, DELIVERABLES AND MILESTONES

D4.2(a) focuses on developing knowledge models that represent the structure and relationships of SMNs, utilizing extensions of the neuro-symbolic DSS introduced in D3.3. These models provide a holistic view of the entire supply network, encompassing various stakeholders and operational processes. The initiative aligns with Task 4.2, which further elaborates on the several types of knowledge models instrumental in improving supply chain visibility and efficiency.

The key types of supply chain network knowledge models developed in this project include:

1. **Physical Model:** Represents the tangible structure of the SMN, including supplier locations, manufacturing plants, warehouses, and transportation routes. This model integrates inputs from T3.2 and T3.3.
2. **Process Model:** Maps out the workflows involved in procurement, production, inventory management, and logistics. Contributions from T3.4 enhance this model's accuracy and applicability.
3. **Information Flow Model:** Captures the movement of data within the SMN, such as order processing, inventory tracking, production updates, and logistics coordination. This model draws from T3.2 and T3.4.
4. **Decision Model:** Represents critical decision-making mechanisms in SMNs, including forecasting, planning, and risk management. It benefits from inputs provided by T3.4.

By using these models, the Intelligent Manufacturing Custodian (IMC) can enhance supply chain visibility, providing real-time insights into material and goods movement, as well as key performance metrics. Through appropriate extensions of the neuro-symbolic DSS in T3.3, the IMC can act as a powerful decision-enabling tool, streamlining supply chain processes, mitigating risks, and optimizing logistics operations to reduce delays and inefficiencies.

This project underscores the importance of data-driven decision-making in SMNs and aims to establish a robust foundation for intelligent supply chain management through advanced knowledge modelling.

2. ARCHITECTURE

This section provides a comprehensive overview of the system architecture, detailing its core component and their interrelations. The architecture is designed to ensure modularity, scalability, and efficiency, with each component playing a distinct yet interconnected role in the overall system. A high-level description of the system’s structure is presented first, outlining the key elements and their interactions. Subsequently, each major component is examined in detail in the following subsections, explaining its functionality, design considerations, and role within the system.

2.1. COMPONENTS AND RELATIONSHIPS

The architecture presented in Figure 1 represents the foundation of the NARRATE project, embodying a comprehensive and flexible framework designed to enable advanced information retrieval, integration, and response generation. This architecture has been entirely developed within the scope of the project, addressing the need for a robust system capable of combining heterogeneous data sources with intelligent reasoning capabilities. At its core, the IMC acts as the central orchestrator, coordinating multiple agents that specialize in distinct tasks. The API module facilitates the dynamic extraction of external data from APIs, ensuring access to up-to-date information, while the IMC interacts with the NSQA (Neural-Symbolic Question Answering) system to retrieve knowledge from ontology knowledge bases.

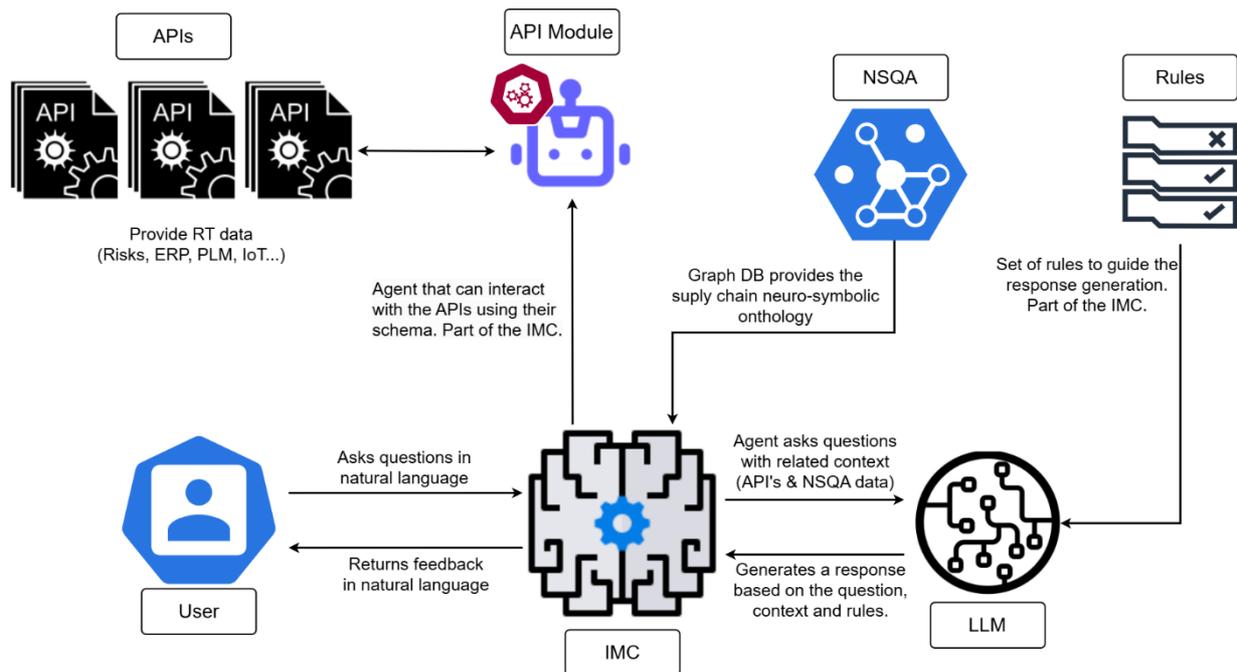


FIGURE 1: ARCHITECTURE DIAGRAM

The IMC governs the entire workflow by defining and applying orchestration rules based on prompt engineering principles, which play a pivotal role in structuring the interactions between the agents and the Large Language Model (LLM). These rules encapsulate domain knowledge, task-specific constraints, and contextual information, guiding the LLM to generate responses that are coherent, explainable, and aligned with the system's objectives. The modular design of the architecture ensures high flexibility and scalability allowing the seamless addition or replacement of agents to adapt to evolving requirements or integrate new data sources. Furthermore, the architecture embraces hybrid reasoning, combining statistical inference from the LLM with symbolic reasoning from the NSQA, which enhances the explainability and reliability of the generated responses.

One of the key innovations of this architecture lies in its capacity to dynamically select and aggregate information from different agents based on the task context, optimizing the quality of the final output. This enables the system to handle a wide range of user queries, from factual questions requiring structured knowledge to more complex requests involving multi-source data synthesis. The design also prioritizes scalability and interoperability, making it suitable for deployment in diverse application domains. Overall, this architecture provides a solid and extensible foundation for intelligent information management, balancing the strengths of LLMs with the precision of symbolic knowledge systems to deliver high-quality, context-aware responses.

2.2. INTELLIGENT MANUFACTURING CUSTODIAN (IMC)

Agent-based orchestration AI systems leverage autonomous agents to manage and coordinate complex tasks across distributed environments, improving efficiency, adaptability, and scalability. These systems are widely used in domains such as cloud computing, smart grids, and robotics, where decentralized decision-making and dynamic task allocation are crucial [1]. By employing multi-agent systems (MAS), these AI-driven orchestrators can handle heterogeneous resources, optimize workflows, and enhance resilience in dynamic environments [2]. Advances in reinforcement learning and game theory further enable agents to make intelligent decisions while interacting with other agents and their environment [3]. The use of agent-based orchestration is becoming increasingly prominent in AI-driven automation, facilitating more efficient and autonomous management of distributed systems.

Agent-based orchestration AI systems are a well-established paradigm, but their state-of-the-art status depends on the specific methodologies and applications used. While traditional MAS and orchestration frameworks have been around for years, recent advancements have pushed the boundaries in several ways:

- **Deep Reinforcement Learning (DRL) in MAS** – New models, such as multi-agent deep Q-networks (MADQN) and actor-critic methods, enable more sophisticated decision-making in complex environments [4].
- **Graph Neural Networks (GNNs) for Agent Coordination** – GNNs are being used to enhance communication and decision-making among agents in dynamic systems [5].
- **Large Language Models (LLMs) for Agent Collaboration** – Integrating LLMs like GPT-4 enables more flexible, natural-language-driven orchestration and planning [6].
- **Game-Theoretic and Auction-Based Resource Allocation** – SOTA orchestration often includes advanced incentive mechanisms, like combinatorial auctions and cooperative game-theoretic models, to optimize multi-agent interactions [7].
- **Edge AI and Federated Learning** – Decentralized orchestration using federated learning improves scalability and privacy, making AI systems more resilient to failures and adaptable to real-time constraints [8].

Thus, while agent-based orchestration remains a fundamental concept, cutting-edge approaches incorporate deep learning, advanced optimization, and emerging AI paradigms to enhance efficiency, autonomy, and intelligence.

2.2.1. AGENT-BASED ORCHESTRATION WITH LARGE LANGUAGE MODELS

Large Language Models (LLMs) [Appendix 1] have been chosen for agent-based orchestration over traditional MAS and reinforcement learning (RL) approaches due to their superior adaptability, reasoning capabilities, and ease of integration. Unlike MAS, which often requires complex rule-based coordination and explicit communication protocols [1], LLMs can dynamically generate responses, infer context, and generalize across various tasks without extensive pre-programmed rules. Similarly, while RL-based agents excel in optimizing specific reward-driven tasks [9], they demand significant training data and

computational resources, making them less practical for rapidly evolving orchestration needs. In contrast, LLMs can process natural language instructions, retrieve relevant knowledge, and perform multi-step reasoning, offering a more scalable and intuitive solution.

The article "Understand the LLM Agent Orchestration" (see Figure 2 and [10]) provides an in-depth exploration of constructing agents based on Large Language Models. It introduces a unified framework comprising four key modules:

- **Profile Module:** This module defines and manages the agent's characteristics and behaviours, including roles, goals, abilities, and knowledge. It outlines three methods for generating agent profiles:
 - LLM Generation Method: Automatically creates agent attributes using LLMs.
 - Dataset Alignment Method: Derives agent profiles from real-world data to enhance realism.
 - Combination Method: Merges real data with LLM-generated profiles to balance authenticity and scalability.
- **Memory Module:** This component stores and organizes information from the environment to guide future actions. It consists of short-term and long-term memory, with mechanisms for reading, writing, and reflecting on information.
- **Planning Module:** Assists agents in decomposing complex tasks into manageable sub-tasks and formulating effective strategies. It distinguishes between feedback-independent planning, which does not rely on post-task feedback, and feedback-based planning, which adjusts plans based on execution outcomes.
- **Action Module:** Transforms abstract decisions into concrete actions, bridging the agent's internal processes with the external environment. Actions are purposeful and can be generated by consulting past experiences or following preset plans.

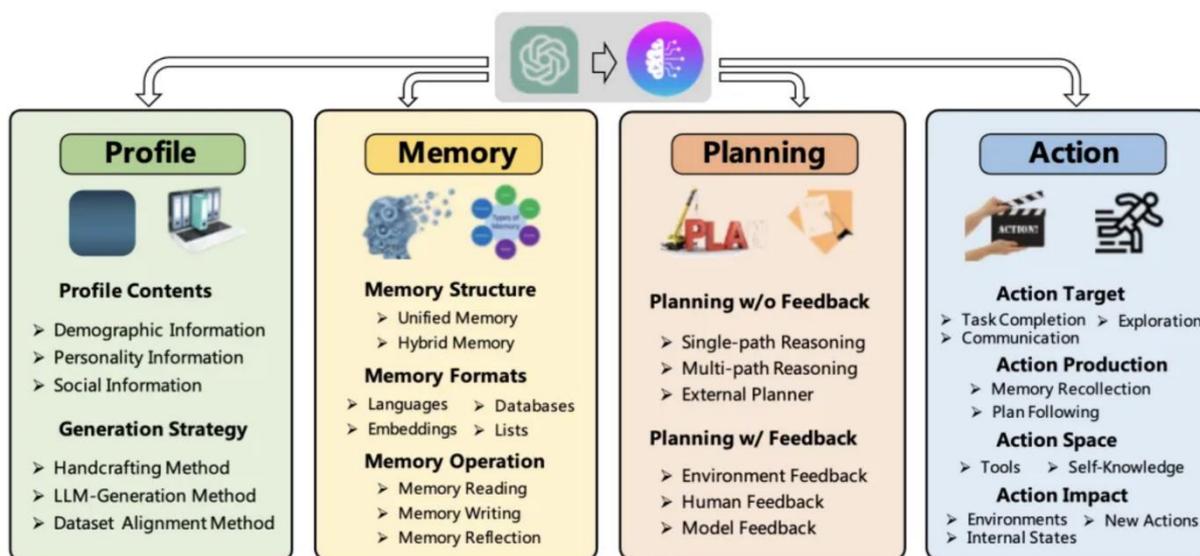


FIGURE 2: LLM AGENT ORCHESTRATION.

To implement this approach effectively, LangChain, a Python framework designed for building context-aware and reasoning-driven applications with LLMs, can be utilized [11]. LangChain provides essential components such as memory, tool integration, and agent architectures, enabling the creation of modular and intelligent orchestrators that interact seamlessly with APIs, databases, and real-world applications. Python’s rich ecosystem further enhances this implementation, with libraries such as OpenAI’s API for state-of-the-art models [12] and FAISS for efficient vector-based retrieval [13]. This combination allows LLM-powered agents to perform dynamic decision-making, manage structured and unstructured data,

and continuously improve through iterative feedback loops. By using LLMs with LangChain, a more flexible, efficient, and scalable orchestration system is achieved, capable of adapting to diverse operational needs.

2.3. NEURO-SYMBOLIC QUESTION ANSWERING (NSQA)

Neuro-Symbolic Question Answering (NSQA) is a powerful approach that blends neural networks with symbolic reasoning to enhance the interpretation and answer extraction from structured knowledge bases. A key element of NSQA systems is the use of ontologies—formal representations of knowledge in a particular domain. Ontologies define a set of concepts and the relationships between them, enabling more meaningful query interpretation.

When these ontologies are stored and queried within graph databases, like Neo4j, the ability to represent complex, interconnected data in a natural, flexible way significantly enhances the system’s reasoning capabilities. Graph databases excel in situations where relationships between entities (e.g., components in a manufacturing system or entities in a business model) are critical to answering queries. By structuring knowledge as nodes (entities) and edges (relationships), graph databases allow NSQA systems to perform more advanced, context-aware reasoning, facilitating the generation of accurate and relevant answers from a knowledge base [14].

In manufacturing, ontologies play a vital role by organizing domain-specific knowledge into a standardized structure. These ontologies encapsulate the interrelationships between various manufacturing components, such as machines, processes, raw materials, and products. By structuring knowledge in this way, manufacturers can enhance decision-making, improve operational efficiency, and ensure data interoperability across different systems.

For example, a manufacturing ontology might define relationships such as "performs," "produces," and "requires," which describe the interactions between machines, processes, and materials, see Figure 3. This structure allows manufacturers to quickly query and retrieve relevant information about how machines perform specific tasks, what materials are required for each process, and how different stages of production are interconnected. Additionally, ontologies in manufacturing can assist in integrating data from disparate systems, making it easier to manage operations and maintain flexibility as processes evolve [15].

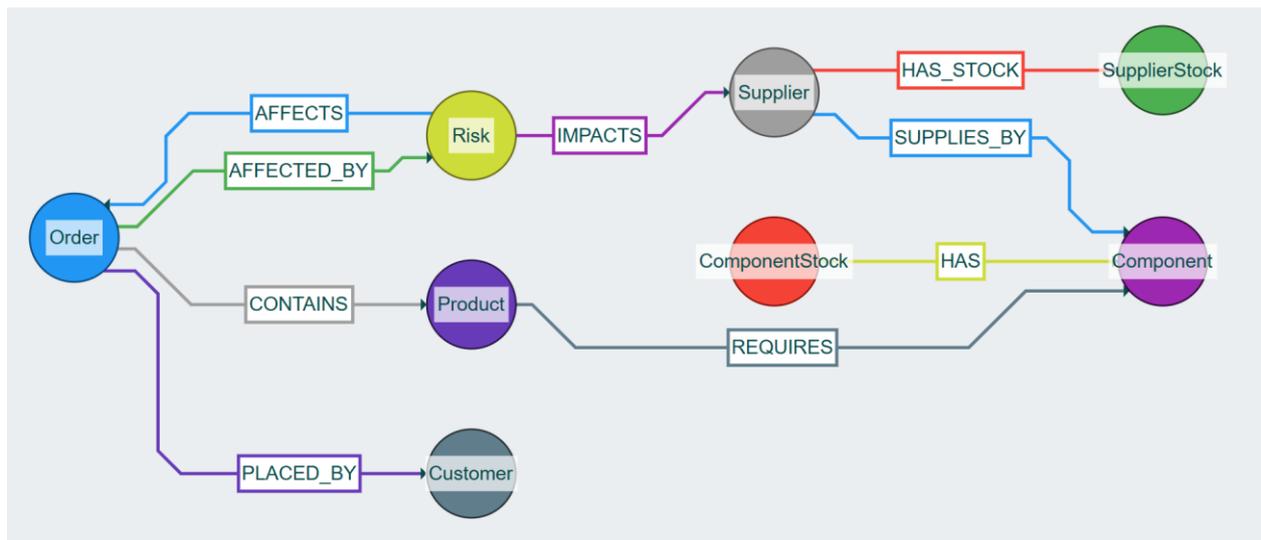


FIGURE 3: GRAPH ONTOLOGY FOR SUPPLY CHAIN EXAMPLE

A review by Sapel and Comet (2024) classifies manufacturing ontologies into various categories, highlighting their different applications within the industry. Their study emphasizes that ontologies are

crucial for enhancing system interoperability and ensuring seamless data exchange between different manufacturing processes and software systems. Using these classifications, manufacturers can adopt the most suitable ontology frameworks to improve their production workflows [16].

According to Fraga and Vegetti (2024), ontology-based solutions significantly improve interoperability among product lifecycle management (PLM) systems by providing standardized representations of manufacturing data. This ensures that different software systems within the manufacturing pipeline can seamlessly exchange information, reducing inefficiencies and improving coordination. The integration of Neo4j within such ontology-based frameworks allows for optimized querying and relationship mapping, enhancing overall system performance [17].

By implementing ontologies, manufacturers can bridge gaps between different information silos, enabling seamless communication across various departments and enhancing the quality of decisions based on comprehensive, structured knowledge. Fraga and Vegetti (2024) also emphasize the importance of ontology-based solutions for achieving interoperability among product lifecycle management systems, ensuring smooth data exchange and improved coordination in manufacturing operations.

2.3.1. ONTOLOGY WEB LANGUAGE (OWL)

The Ontology Web Language (OWL) is a semantic web language specifically designed to define, structure, and share ontologies across different systems and platforms. It enables the representation of complex knowledge in a machine-readable format, making it a powerful tool for knowledge management and data interoperability in various domains, including manufacturing.

One of the primary reasons OWL is suitable for developing manufacturing ontologies is its strong formal semantics. Unlike simple data models or taxonomies, OWL provides a more expressive framework that allows for detailed classifications, logical constraints, and relationships between different entities in the manufacturing domain. With these capabilities, manufacturers can define intricate dependencies between various components, processes, and resources, ensuring a more accurate representation of real-world scenarios.

Furthermore, OWL supports automated reasoning, which is a key advantage in manufacturing systems. Reasoning mechanisms enable systems to infer new knowledge from existing data, detect inconsistencies, and validate logical relationships within the ontology. This means that manufacturers can utilize OWL not just for data organization, but also for advanced decision-making, process optimization, and troubleshooting potential inefficiencies in production workflows.

Another significant feature of OWL is its compatibility with existing web technologies and standards. Since modern manufacturing systems increasingly rely on digital transformation, Industry 4.0 initiatives, and Internet of Things (IoT) integrations, having an ontology language that aligns with widely adopted web standards is crucial. OWL's ability to integrate seamlessly with other semantic web technologies, such as the Resource Description Framework (RDF) and the SPARQL query language, enhances its utility in distributed and networked manufacturing environments.

By using OWL to develop manufacturing ontologies, organizations can ensure interoperability across different software platforms, improve data sharing between stakeholders, and create a more unified framework for representing manufacturing knowledge. This, in turn, leads to better collaboration, streamlined automation, and enhanced efficiency across the entire production lifecycle.

Thus, Lemaignan, S., & Siadat, M. argue that OWL is not only a suitable choice for manufacturing ontologies but also a strategic tool for advancing intelligent manufacturing systems. By adopting OWL-based ontologies, manufacturers can facilitate seamless communication between different applications, reduce data silos, and enable more sophisticated analytics for continuous improvement in industrial operations [14].

2.3.2. NEO4J GRAPH DATABASE

Neo4j, a widely used graph database, provides an ideal platform for implementing manufacturing ontologies. Its graph-based data model aligns perfectly with the structure of ontologies, where entities (such as machines, products, or processes) are represented as nodes and relationships between them (such as "performs" or "requires") are depicted as edges. This natural mapping of domain knowledge to a graph structure allows for efficient representation and querying of complex interconnected data.

Neo4j offers several key advantages when using it for ontology management in manufacturing:

- **Flexibility and Scalability:** Neo4j's graph structure allows manufacturers to model a wide variety of data relationships, from basic machine-process interactions to more complex multi-step workflows. The flexibility of the graph model also allows for easy scalability as the manufacturing system grows and evolves.
- **Efficient Querying:** The Cypher query language used by Neo4j enables users to efficiently query highly connected data. For example, querying which machines perform certain processes or which materials are required for specific production steps is straightforward and intuitive with Cypher.
- **Integration and Real-Time Analytics:** Neo4j's ability to integrate with other data sources and its support for real-time querying are particularly valuable for adaptive manufacturing environments. As production lines change or new data sources become available, Neo4j can dynamically incorporate these changes, ensuring the system remains responsive and up to date.
- **Graph Traversal:** One of the standout features of Neo4j is its ability to traverse large, complex graphs quickly, enabling the identification of dependencies and relationships within the manufacturing ecosystem. For example, it can trace material flows, determine the optimal machines for a specific task, or identify bottlenecks in the process [\[14\]](#).

Neo4j's suitability for these tasks makes it a valuable tool in the manufacturing sector, where efficient data management and real-time insights are key to operational success. With Neo4j, manufacturers can create dynamic, connected data models that support intelligent decision-making, predictive maintenance, and production optimization.

In a typical graph database structure for manufacturing ontologies, various components—such as machines, processes, products, and materials—are represented as nodes. These nodes are connected through edges that define their relationships. For instance, a "machine" node could be linked to a "process" node through an edge labelled "performs," indicating that the machine is responsible for executing a particular process. Similarly, a "process" node could be connected to a "material" node via an edge labelled "requires," signifying that the process needs specific raw materials to be carried out.

This structure allows for a variety of insightful queries. For example, a manufacturer might query which processes a particular machine can perform or which materials are used in the production of a specific product. By navigating the graph, one can also determine which products require a certain process or which machines are associated with particular production stages [\[Appendix 2\]](#).

Additionally, graph databases like Neo4j make it easy to adapt and modify this structure as the manufacturing system evolves. New relationships can be introduced as required (e.g., new types of machines, processes, or materials), and the database can scale with the increasing complexity of the manufacturing ecosystem.

The inherent flexibility of graph databases allows for advanced analytics, where manufacturers can perform in-depth queries and gain insights into their operations, such as identifying production inefficiencies or predicting maintenance needs based on historical data [\[15\]](#).

2.4. API CONNECTION MODULE

As industries continue to embrace data-driven transformation, the ability to manage, integrate, and utilize information across complex networks has become increasingly vital. Ensuring seamless data interoperability, enhancing contextual understanding, and enabling intelligent decision-making are key challenges in modern supply chain and manufacturing ecosystems. Advancements in data modelling, digital representations, and standardized communication frameworks are driving innovative approaches to these challenges, fostering more efficient and connected operations. By adopting structured methodologies and innovative technologies, organizations can enhance visibility, optimize processes, and unlock new opportunities for innovation. This section explores foundational concepts and methodologies that support these objectives, highlighting their role in shaping the future of data-driven industrial ecosystems.

2.4.1. KNOWLEDGE MODELS IN A HYBRID FEDERATED DATA MESH (D3.1)

In modern industrial ecosystems, data interoperability and contextual understanding are key challenges in enabling seamless information exchange across supply chain networks. Knowledge models serve as structured frameworks that organize, annotate, and enhance datasets, ensuring data standardization, discoverability, and usability. By integrating Blueprint Frames and a Hybrid Federated Data Mesh, organizations can improve data management, enhance decision-making, and foster collaborative intelligence across a decentralized supply chain environment.

Blueprint Frames as the Foundation for Knowledge Models

Blueprint Frames represent modular, structured digital entities that encapsulate the key elements of a Smart Manufacturing Network (SMN). These frames are built upon semantic data models, ensuring alignment with Product Lifecycle Management (PLM), Enterprise Resource Planning (ERP), IoT sensors, historical production records, and industry standards. Key types of Blueprint Frames include:

- **Supplier Data Frame** – Captures real-time supplier performance, lead times, risk factors, and sustainability practices.
- **Product Frame** – Defines product specifications, material compositions, and design tolerances for quality assurance.
- **Production Frame** – Maps the step-by-step manufacturing processes, machine utilization, and quality control measures.
- **Sensor Data Frame** – Aggregates real-time IoT sensor readings for predictive maintenance and process optimization.
- **Logistics Frame** – Tracks inventory levels, transportation routes, and supply chain disruptions.
- **Sustainability Frame** – Monitors energy efficiency, carbon footprints, and waste management metrics.

By incorporating Blueprint Frames into a Hybrid Federated Data Mesh, organizations can dynamically integrate and contextualize multiple datasets from various sources, such as supplier networks, material flows, production lines, and logistics providers. This approach enhances visibility, adaptability, and decision intelligence across distributed networks.

Enabling Federated Data Interoperability

A Hybrid Federated Data Mesh enables seamless integration of diverse, decentralized datasets by establishing a unified virtual federated database. Key advantages include:

- **Standardized Data Exchange** – Ensures cross-platform compatibility through ontology-based representations and OpenAPI frameworks.

- **AI-Driven Insights** – Leverages machine learning models to identify bottlenecks, predict risks, and optimize decision-making.
- **Automated Contextualization** – Dynamically annotates and categorizes data based on real-time events and operational conditions.
- **Resilient Supply Chain Orchestration** – Supports proactive risk management and adaptive logistics planning.

By implementing these methodologies, organizations enhance data transparency, drive automation, and unlock new opportunities for innovation, ultimately shaping the future of data-driven industrial ecosystems.

Note that this information is just summarized here for better comprehension. For a deeper understanding, please refer to deliverable 3.1.

2.4.2. DIGITAL TWIN REPRESENTATION IN MANUFACTURING NETWORKS (D3.2)

The adoption of Digital Twins in manufacturing networks represents a paradigm shift toward intelligent, adaptive, and self-orchestrating production ecosystems. Digital Twins provide real-time virtual representations of products, machines, processes, logistics, and environmental factors, enabling manufacturers to optimize operations, predict disruptions, and streamline supply chain coordination.

To maintain a clear separation of concerns, Digital Twins in manufacturing networks are categorized into six primary domains:

- **Product Digital Twin:** Simulates product lifecycles, materials, and design tolerances.
- **Machine Digital Twin:** Models individual machines, their operational efficiency, and predictive maintenance needs.
- **Process Digital Twin:** Represents workflow sequences, resource allocations, and production optimization.
- **Sensor Digital Twin:** Integrates IoT sensor data for real-time monitoring and anomaly detection.
- **Logistics Digital Twin:** Provides visibility into transportation networks, warehouse management, and inventory flow.
- **Environmental Digital Twin:** Assesses sustainability metrics, energy consumption, and ecological impact.

By integrating these domain-specific twins into a SMN, manufacturers gain holistic visibility, operational intelligence, and real-time adaptability.

Blueprint Frames: The Core Structure of Digital Twins

Blueprint Frames serve as the data foundation for Digital Twins, structuring and interconnecting critical elements within a manufacturing network. They enable:

1. **Virtual Modelling of Manufacturing Systems:** Digital Twins use Blueprint Frames to create real-time simulations of production lines, allowing manufacturers to test, refine, and validate various scenarios before implementation.
2. **Adaptive & Programmable Manufacturing:** By incorporating Digital Twin Processing Language (DTPL), manufacturers can dynamically adjust production workflows in response to supply chain fluctuations, demand surges, or unexpected disruptions.

3. **AI-Driven Predictive Analytics:** Digital Twins leverage machine learning models to monitor and analyse KPIs, identify inefficiencies, and recommend process optimizations in material usage, logistics, and energy consumption.
4. **Seamless Interoperability Across Factories & Suppliers:** Standardized Blueprint Frames enable multi-site synchronization, ensuring seamless data exchange between factories, suppliers, and logistics providers.
5. **Proactive Risk Mitigation:** Continuous updates to Blueprint Frames allow Digital Twins to anticipate and model alternative production scenarios, ensuring resiliency against supply chain disruptions.

DTPL: Enabling Intelligent Automation in Digital Twins

The Digital Twin Processing Language provides a programmable framework for managing and orchestrating digital twin interactions. It introduces event-driven processing, AI-powered analytics, and workflow automation, transforming static manufacturing models into dynamic, self-adjusting networks.

Key capabilities of DTPL include:

- **Semantic Awareness & Interoperability:** Facilitates structured data exchange across Digital Twin ecosystems.
- **Event-Driven Decision-Making:** Automatically responds to disruptions (e.g., supplier delays, machine failures) by triggering corrective actions.
- **Predictive & Prescriptive Analytics:** Integrates AI models to forecast potential disruptions and recommend optimal responses.
- **Resource Optimization:** Dynamically reallocates resources based on real-time demand fluctuations and machine availability.
- **Self-Healing Systems:** Enables manufacturing lines to detect, isolate, and mitigate failures automatically.
- **Sustainability Optimization:** Reduces energy consumption through intelligent power allocation and demand forecasting.

By integrating Blueprint Frames, Digital Twins, and DTPL, manufacturers transform supply chain operations into adaptive, self-regulating systems. This enables a fully interconnected, real-time decision-making framework that enhances efficiency, minimizes risks, and drives industrial innovation.

Note that this information is just summarized here for better comprehension. For a deeper understanding, please refer to deliverable 3.2.

2.4.3. INTEGRATION OF LANGCHAIN AND OPENAPI STANDARDS

The OpenAPI Specification (OAS), formerly known as Swagger, is a widely adopted standard for describing RESTful APIs in a machine-readable format. OpenAPI allows developers to define the structure of API endpoints, including the request and response formats, authentication requirements, and error handling. This standardized approach makes APIs more discoverable, and interoperable, and can be used to generate client libraries, documentation, and even mock servers [\[18\]](#).

Using APIs that adhere to the OAS is the optimal approach for enabling IMC agents to efficiently query and interact with data from the DTs. This standardized framework offers multiple advantages that enhance accessibility, automation, and interoperability. The main advantages of using OAS for the API module can be described as:

- **Consistency and Accessibility:** The OAS establishes a clear, standardized way for both humans and systems to interpret and interact with APIs, reducing ambiguity and simplifying integration. This ensures that API consumers can understand available endpoints, expected inputs, and outputs without requiring direct access to the source code [18].
- **Seamless Agent Communication:** By providing a structured framework, OpenAPI enables IMC agents to efficiently locate, request, and process relevant data with minimal configuration. The standardized API definitions help agents dynamically adapt to available resources and retrieve actionable information as needed [19].
- **Automation and Efficiency:** OpenAPI facilitates the automatic generation of API documentation and client libraries, reducing development effort and ensuring consistency across different implementations. This automation enhances maintainability and accelerates the integration process [20].

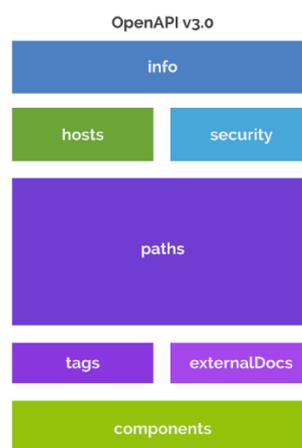


FIGURE 4: OPENAPI V3.0 STRUCTURE. SOURCE: <https://www.apideck.com/blog/introduction-to-openapi-specification>

Incorporating LangChain and OpenAPI standards enables the creation of intelligent agents capable of interacting with APIs. The agent's role in this process involves analysing user input, determining the appropriate API to interact with, and constructing the necessary requests. OpenAPI acts as a structured guide for the agent, outlining the available actions, request types, and response formats, allowing the agent to seamlessly understand and navigate the API.

By using LangChain and OpenAPI, an agent can:

- Parse OpenAPI documentation to identify which endpoints are relevant to the task at hand [21].
- Automatically generate API requests based on the user's input, following the specifications laid out in the OpenAPI definition [21].
- Handle various response formats, such as JSON or XML, and process the results to present them in an understandable manner to the user [21].
- Implement logic to handle error responses, missing data, or exceptions, ensuring the agent can respond appropriately based on the given context [21].

When implementing an agent to interact with APIs using OpenAPI documentation, the following steps are essential:

First, the OpenAPI specification, available in formats like JSON or YAML, is loaded into the LangChain agent. This serves as the primary reference for understanding the structure of the API, including its endpoints and the required parameters. Next, LangChain tools are created that correspond to specific API

endpoints outlined in the OpenAPI specification. These tools are responsible for executing the API calls, handling HTTP requests, and processing the responses.

The agent itself is then defined with the task of interacting with the user. Based on the user's input, the agent determines which tool to use. For instance, if the user requests weather information, the agent will invoke the weather API tool, which was defined based on the OpenAPI documentation.

As the user input is received, the agent dynamically generates the necessary API requests, referring to the OpenAPI specification to construct query parameters or send data via POST requests. Once the response from the API is received, the agent processes the data according to the defined logic. This may include transforming raw data into a more user-friendly format or handling any errors that occur, such as missing information or failed API calls.

In this way, the combination of LangChain and OpenAPI enables the development of intelligent agents capable of interacting with a variety of external APIs in an efficient and structured manner.

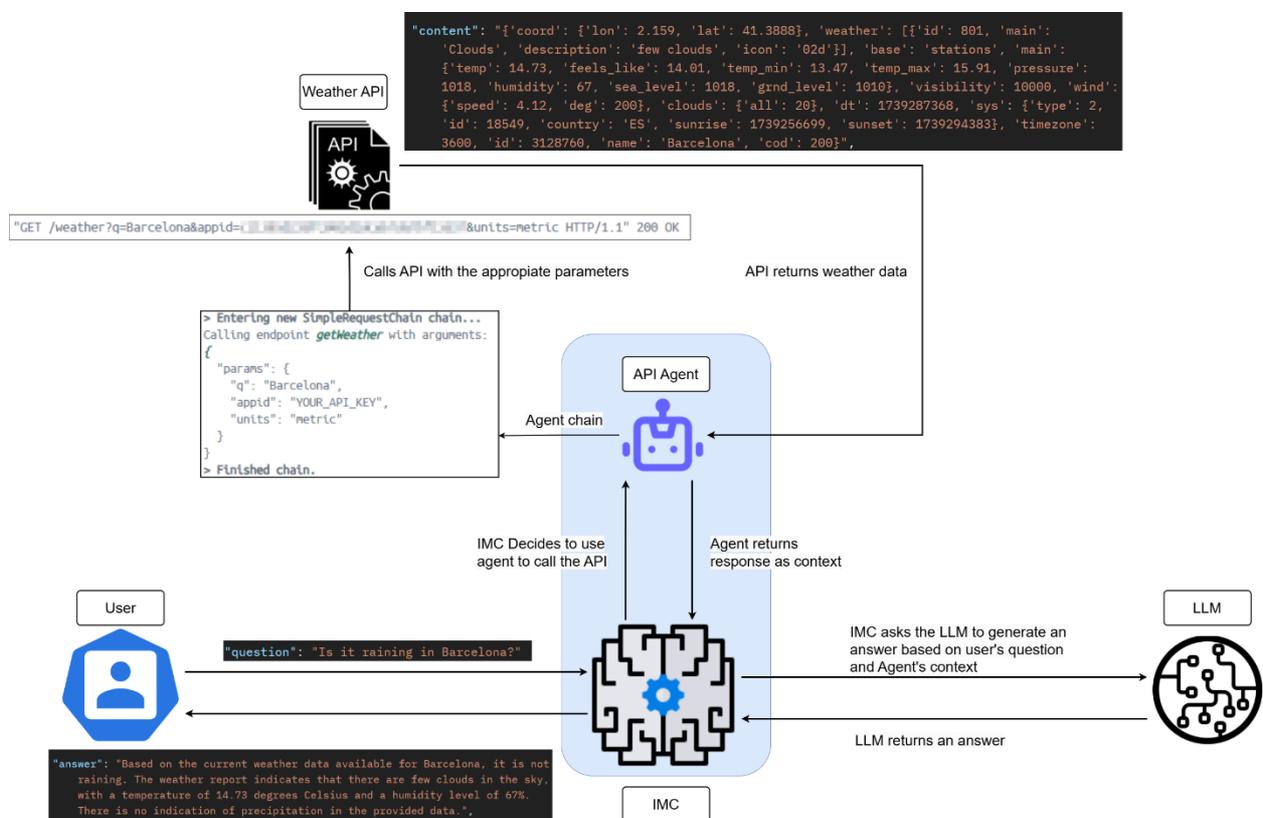


FIGURE 5: EXAMPLE OF IMC USING AN API TO RETRIEVE WEATHER DATA

2.4.4. API-DRIVEN INTEGRATION OF TOOLS WITHIN THE IMC

To ensure seamless interoperability and efficient data exchange within the Intelligent Manufacturing Control (IMC) Platform, we propose that all tools listed in the Tools & Owners Table provide data to the IMC using APIs (Application Programming Interfaces), adhering to the OpenAPI standard. This approach guarantees:

- Standardized Data Exchange: Ensuring uniform communication between diverse tools, Digital Twins, AI models, and data sources.

- Scalability & Extensibility: Allowing new tools to be easily integrated into the IMC ecosystem without extensive modifications.
- Real-Time Interoperability: Enabling continuous data updates and automated decision-making across the AI-driven network.
- Secure & Controlled Access: Ensuring authentication, encryption, and role-based access for data exchange.

2.4.5. TOOLS AND PROPOSED API INTEGRATION

Considering the architectural framework outlined in D3.1 Design and implementation of contextualization services (a) and D3.2 Digital Twin design and development (a), along with the tools described in WP6 Intelligent Manufacturing Custodian and AI Platform: implementation and assessment, and leveraging the capabilities of LangChain and the OpenAPI standard for creating intelligent agents, we propose that the following tools be made accessible via API while adhering to OpenAPI standards.

By implementing this approach, the API module can act as an intelligent agent that retrieves relevant data from each tool based on specific information requirements. Since LangChain utilizes the OpenAPI standard to generate agents, ensuring API accessibility through this standard will enable seamless integration, dynamic agent creation, and efficient interaction between different components of the system. This approach enhances modularity and flexibility while streamlining data retrieval and processing.

With this approach, the IMC would utilize the API module to retrieve data from a specific source based on the information required to respond to a user's query. The intelligent agent would autonomously determine which endpoints need to be called to obtain the necessary data, using the OpenAPI standard to facilitate this process automatically while remaining resilient to changes. This means that the API implementation of each tool can evolve, but as long as these changes are accurately reflected in the schema, the API module will continue to interact with them seamlessly.

Below, you will find a list of the tools in WP6 and how they could be integrated using an API.

D2.1 - Resilience Strategy and Tool (a) (BUL)

- **Function:** Stores and codifies response plans for risk classification and impact analysis.
- **API Integration:** Exposes risk classification results, impact assessments, and response strategies via an OpenAPI endpoint for IMC retrieval.

D2.2 - Risk Identification & Monitoring Tool (SAG)

- **Function:** A Complex Event Processing (CEP) tool that provides real-time risk identification and dashboards.
- **API Integration:** Provides real-time risk alerts and monitoring data through webhook-based API updates for immediate IMC processing.

D2.3 - Supplier & SMN Risk Assessment Tool (BUL)

- **Function:** Assesses risk in supplier networks and the Smart Manufacturing Network (SMN).
- **API Integration:** Connects to the Risk Blueprint Plugin, feeding supplier risk ratings and SMN vulnerabilities into the IMC via REST APIs.

D3.1 - Design and implementation of Contextualization Services (a) (SERV)

- **Function:** Establishes semantic context between data sources (ERPs, IoT, DTs) and computational models (Spark).

- **API Integration:** Provides real-time contextualized data pipelines through GraphQL APIs, facilitating AI-powered decision-making.

D3.2 - Digital Twin design and development (a) (SERV)

- **Function:** Acts as the AI Platform's central Digital Twin to model, simulate, and predict manufacturing processes.
- **API Integration:** Serves real-time DT simulation outputs through a bi-directional API, allowing the IMC to request and push data dynamically.

D3.3 - SMN Knowledge Model using Neuro-symbolic DSS (a) (NUN)

- **Function:** Uses Neuro-symbolic Decision Support Systems (DSS) for knowledge representation.
- **API Integration:** Offers a knowledge graph query API, enabling IMC access to inference-based decision recommendations.

D3.4 - Automated Workflows & Process Orchestration (a) (IPK)

- **Function:** Automates and orchestrates AI-driven manufacturing workflows.
- **API Integration:** Provides workflow automation triggers via event-driven APIs, ensuring real-time process optimization.

D3.5 - Security & Privacy Services (INSA)

- **Function:** Ensures secure access, authentication, and compliance.
- **API Integration:** Implements an OAuth 2.0 API for authentication and secure data exchange within IMC transactions.

D4.1 - Resilience, Sustainability & Circularity Stress Testing Tool (a) (SAG)

- **Function:** Evaluates system stress responses under operational uncertainties.
- **API Integration:** Exposes stress test results and sustainability impact scores through an API, feeding IMC's risk management models.

D4.2 - End-to-End AI-driven Visibility Model & Support DSS (a) (NUN)

- **Function:** Provides a holistic view of manufacturing processes.
- **API Integration:** Supplies real-time operational KPIs via an event-driven API, supporting IMC's AI-based analytics.

D4.3 - Production Planning & Process Routing System and Algorithms (a) (INSA)

- **Function:** Optimizes production scheduling and process routing.
- **API Integration:** Offers a planning API that enables IMC to dynamically adjust scheduling in response to changing conditions.

D4.4 - Intelligent Logistics & Warehousing Modules (a) (DHL)

- **Function:** Manages logistics, inventory, and warehousing operations.
- **API Integration:** Provides an inventory tracking API and logistics updates for IMC integration.

D4.5 - Reconfiguration of Production (INSA)

- **Function:** Enables adaptive production reconfiguration.
- **API Integration:** Exposes adaptive reconfiguration decisions via a RESTful API, allowing IMC to modify production workflows in real time.

2.4.6. API INTEROPERABILITY FRAMEWORK

To unify these APIs within the IMC platform, we propose a three-layered API architecture:

1. **Data Ingestion APIs:** Tools push real-time data into the IMC via secure RESTful, GraphQL, or WebSocket APIs.
2. **Decision & Computation APIs:** AI models, Digital Twins, and DSS systems exchange insights with the IMC through an OpenAPI-compliant event-driven framework.
3. **Orchestration & Execution APIs:** The IMC triggers workflow actions in connected tools through callback APIs and microservices architecture.

By using OpenAPI standards, event-driven architectures, and AI-enhanced interoperability, we ensure that all tools seamlessly integrate with the IMC, driving real-time, data-driven decision-making in the SMN.

2.5. RULES MODULE

The Rules Module is a fundamental component in agent-based orchestration systems that utilize Large Language Models. This module defines a structured set of rules, formulated in natural language, which serve as explicit constraints and guidelines for LLM-driven agents. Using prompt engineering techniques, the Rules Module enhances consistency, interpretability, and adaptability in agent decision-making processes.

2.5.1. THE ROLE OF THE RULES MODULE IN LLM-BASED ORCHESTRATION

The Rules Module provides a flexible and easily modifiable framework, unlike traditional MAS that rely on hardcoded rules or RL-based approaches that require extensive training data [22]. These rules act as dynamic prompts that guide LLMs in reasoning, decision-making, and action execution. The structured nature of these rules ensures that the agents operate within defined parameters, improving alignment with system goals and reducing the likelihood of undesired behaviour.

The Rules Module supports various types of rules, including:

- **Normative Rules:** Define acceptable behaviour for agents in different scenarios.
- **Constraint-Based Rules:** Impose operational limits, such as resource constraints or access controls.
- **Priority Rules:** Help agents prioritize tasks based on urgency or importance.
- **Safety and Compliance Rules:** Ensure that agent actions adhere to ethical and legal considerations.

By employing these structured rule sets as prompt templates, LLM-powered agents can dynamically interpret and adapt their behaviour without requiring extensive retraining [23].

2.5.2. ENHANCING AGENT-BASED ORCHESTRATION WITH STRUCTURED PROMPTS

Prompt engineering plays a crucial role in designing effective rule-based interactions for LLM-driven agents. It involves crafting precise and context-aware instructions that optimize LLM responses. Well-

designed prompts improve response accuracy, reduce ambiguity, and enhance the reliability of AI-driven orchestration.

Several best practices in prompt engineering contribute to the effectiveness of the Rules Module:

- **Few-Shot and Zero-Shot Learning:** By structuring rules as example-driven prompts, agents can generalize knowledge efficiently without requiring large, labelled datasets [24].
- **Contextual Embedding:** Rules can be embedded within hierarchical contexts, allowing agents to differentiate between general guidelines and task-specific instructions [25].
- **Iterative Refinement:** Dynamic feedback loops enable continuous improvement of prompts, ensuring that the rules remain relevant as operational requirements evolve [26].

Compared to traditional rule-based systems or predefined policy-driven models, prompt engineering provides a more adaptive and scalable solution. Its advantages include:

- **Flexibility:** Rules can be modified or expanded without retraining models, reducing development overhead.
- **Interpretability:** Natural language rules improve transparency, making it easier for stakeholders to understand and refine system behaviour.
- **Cross-Domain Adaptability:** LLMs using prompt-based rule sets can operate across diverse domains, from cloud computing to robotic automation, without requiring domain-specific retraining [27].

Recent research highlights how prompt engineering significantly enhances the performance of LLM-driven orchestration. Studies have shown that structured prompts improve task efficiency and reduce hallucination in generative models [28]. Furthermore, frameworks like LangChain provide advanced tools for integrating rule-based prompts, memory management, and external data retrieval, further optimizing agent-based orchestration [29].

3. THEORETICAL PROOF OF CONCEPT

In the explanation below, we will walk through a simple, complete flow of the system to illustrate how the different steps work together seamlessly. For clarity, the example will be kept basic, focusing on a straightforward process to highlight the core functionality. By using a simplified approach, we aim to demonstrate the key concepts and how the system progresses from end-to-end, ensuring a clear understanding of how the overall flow operates without getting into complex details.

3.1. DESCRIPTION OF MODULES CONFIGURATIONS

For this simple example, which demonstrates a complete flow, we will outline the configurations of each module and describe their behaviour. Below, we will explain each module's configuration.

- IMC: As discussed in section 2.2, Large Language Models have been chosen for agent-based orchestration. The IMC will contain an orchestrator agent that will create the plan to follow, a graph agent to make queries to the NSQA module, an API agent that will get Information from the digital twins by making HTTP request to certain endpoints, and a question answering agent that will generate a final response based on the contextual information provided by the different sources.
- NSQA: The ontology depicted in figure 6 represents a simple structured model for managing supply chain relationships, emphasizing the flow of information between key entities.

At its core, an **Order** is placed by a **Customer** and contains one or more **Products**. Each product is associated with specific **Components**, which are essential for manufacturing a certain type of product. The availability of these components is managed through **ComponentStock**, ensuring sufficient inventory. Components are sourced from **Suppliers**, who maintain their own **SupplierStock**. The relationships between these entities, such as "REQUIRES" for product dependencies and "SUPPLIES_BY" for supplier contributions, ensure seamless supply chain visibility and traceability.

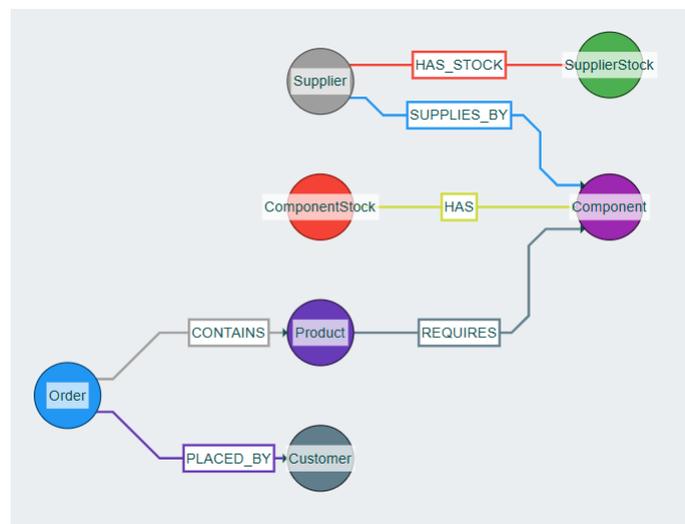


FIGURE 6: EXAMPLE OF GRAPH DATABASE USED FOR THE PROOF OF CONCEPT

- API: The API module will only have access to a single risk API for this proof of concept, and it will be used to provide insights into risks associated with suppliers. For example, it will include climatological risks that could impact a supplier's ability to deliver sufficient stock. If a specific supplier operates in a region prone to extreme weather events, such as hurricanes, droughts or torrential rains, the API will provide relevant risk assessments. Note that this is just a simple example to show how the API module would interact with an API.

- Rules: The rules module will be represented by different prompts that help the LLM generate different kinds of response:
 - Example of prompt for graph database:

Use of a cypher query for a graph database based on a set of rules and a schema. The query should only use the specified relationship types and properties in the schema, avoiding any others. Key instructions include:

 - Product names must be in singular form and start with a capital letter.
 - Only data present in the graph should be returned; no additional cases should be created.
 - Relationships, such as [r:REQUIRES], should be included in the query.
 - The assistant is provided with specific example queries and scenarios, such as how to check stock for orders or how to calculate how many products can be built with the available stock.
 - The schema and user query are provided, and the assistant is expected to generate an accurate Cypher query based on those guidelines. It must not offer explanations or apologies, only the query itself. If the assistant cannot construct a query, it must refrain from doing so and provide a response stating that it cannot create the query.

3.2. CONCEPTUAL WORKFLOW: SUPPLIER QUERY RESPONSE PROCESS

This section outlines the theoretical workflow for how the system would respond to a user query regarding suppliers of a specific material (referred to as X material). The workflow demonstrates the integration of various agents within the system to collect, process, and present relevant information to the user.

Workflow Steps:

1. User Inquiry:
 - A user submits a question, such as, "Who can provide [X material]?"
2. System Initialization:
 - Upon receiving the query, the IMC identifies that the question requires multiple types of data (supplier availability and risk data) to generate a response.
 - The IMC creates a response plan that outlines the necessary steps to gather the required information.
3. Data Collection Phase:
 - Step 1: Supplier Data Acquisition
 - The system activates the NSQA agent, which queries the relevant supplier database to identify those who provide the requested material (X).
 - The agent retrieves additional details such as available stock levels for each supplier.
 - Step 2: Risk Assessment Data
 - Simultaneously, the system activates the plugin agent, which gathers risk-related information for each identified supplier.
 - This data could include financial health, reputation, operational risks, or other relevant factors that affect the supplier's reliability.
4. Data Integration:
 - The IMC consolidates the supplier data (names and stock levels) and risk data (associated risks with each supplier) into a unified set of information.
5. Response Generation:
 - The compiled information is sent to the question answering agent, which processes the data and generates a response.

- The question answering agent structures a clear and comprehensive answer to the user, including:
 - A list of suppliers for the requested material.
 - Stock availability for each supplier.
 - A summary of risk factors related to each supplier.
6. User Output:
- The final response is presented to the user in a way that answers the original question effectively, allowing the user to evaluate potential suppliers based on both material availability and associated risks.

Expected Behaviour:

- The system demonstrates seamless interaction between the **NSQA agent** (to retrieve supplier information) and the **plugin agent** (to gather risk data).
- The **question answering agent** processes this integrated data and returns a user-friendly, actionable response.
- The system is capable of handling variations in the query (e.g., different materials or risk profiles) by adjusting the workflow dynamically.

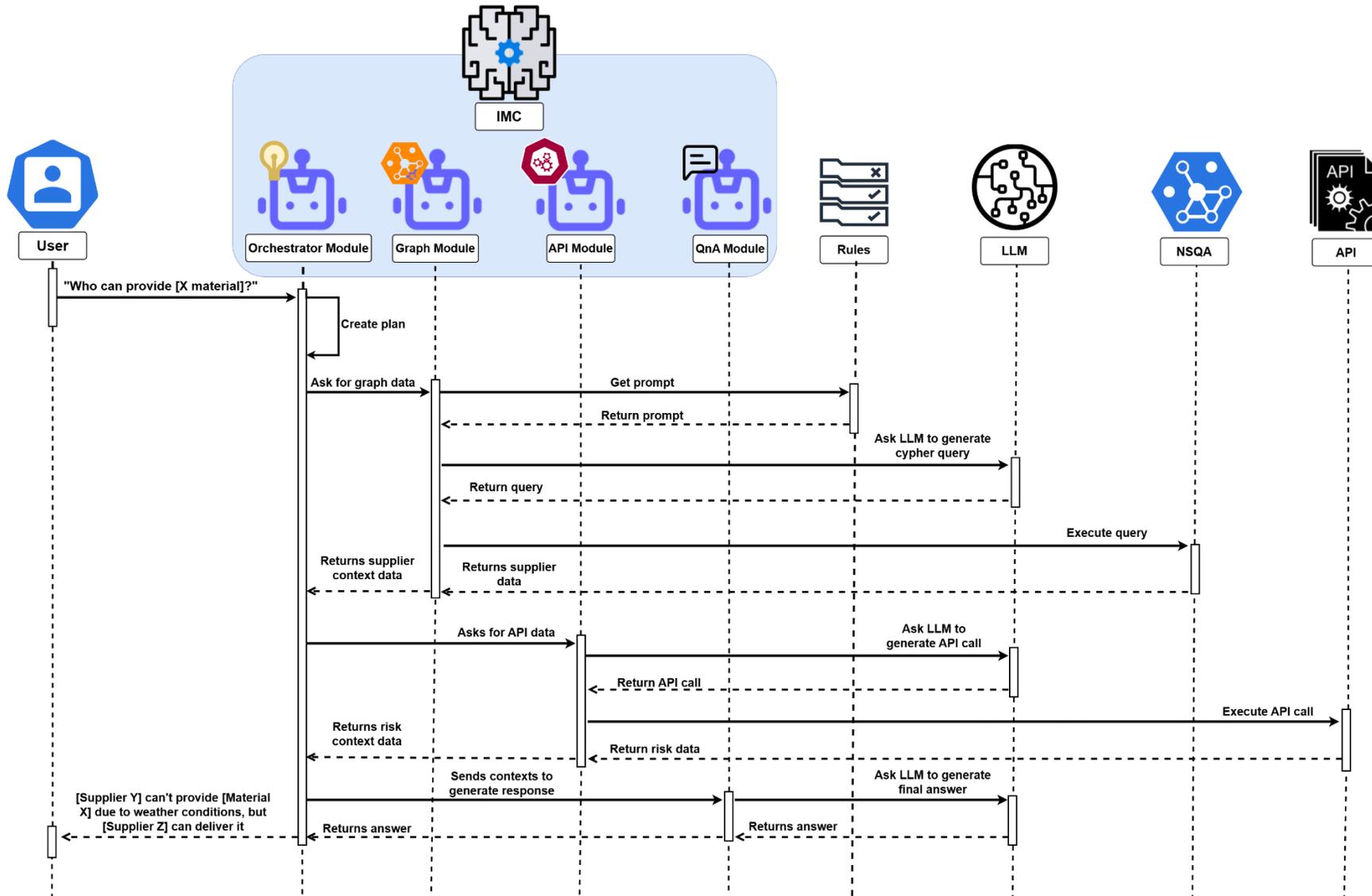


FIGURE 7: PROOF OF CONCEPT SEQUENCE DIAGRAM

4. CONCLUSIONS

The work presented in this deliverable successfully establishes an AI-driven visibility model aimed at improving decision-making and operational efficiency within Smart Manufacturing Networks. By integrating neuro-symbolic reasoning, structured rules, and AI-driven orchestration, this model enhances supply chain transparency, ensuring stakeholders can make informed and timely decisions.

Key findings from this work highlight the significant improvements brought by AI-driven decision support systems, which facilitate better supply chain coordination and optimization. The use of graph-based ontologies has proven instrumental in enhancing data interoperability, providing a structured knowledge representation that enables more effective analysis of supply chain dynamics. Additionally, the structured Rules Module ensures operational consistency, regulatory compliance, and alignment with industry best practices. The API Module further supports seamless integration across various digital platforms, ensuring continuous data flow and interoperability.

Future efforts should focus on refining system interoperability by enhancing data exchange mechanisms and increasing compatibility with emerging industry standards. Further research and development should explore additional AI-driven enhancements to improve the predictive capabilities of the system, enabling proactive decision-making. Real-world pilot implementations will be essential in validating the effectiveness of the proposed model, identifying potential limitations, and refining functionalities to better meet industry demands.

To ensure practical applicability, demonstration implementations of the pilots should be conducted, allowing stakeholders to test and validate the framework under real-world conditions. These demos will help assess the system's robustness, scalability, and adaptability, ensuring that it meets the needs of diverse supply chain environments.

Moreover, scalability considerations should be addressed to ensure that the AI-driven visibility model remains effective in increasingly complex and dynamic supply chain environments. Emphasis should also be placed on continuous advancements in automation, knowledge modelling, and real-time data analytics to further optimize supply chain management and operational resilience.

By advancing these areas, the project can contribute to the long-term transformation of supply chain ecosystems, fostering greater efficiency, adaptability, and intelligence in modern manufacturing operations.

5. APPENDIX

APPENDIX 1: LARGE LANGUAGE MODELS

Large Language Models (LLMs) are a subset of artificial intelligence (AI) designed to understand and generate human language. These models are trained on extensive datasets, enabling them to perform a wide range of tasks, including language translation, text summarization, and code generation. Their versatility has made them integral to various applications, from chatbots to complex data analysis.

Definition and Functionality

At their core, LLMs utilize deep learning techniques, particularly neural networks, to process and generate language. They analyse vast amounts of text data to learn patterns, grammar, and context, allowing them to produce coherent and contextually relevant outputs. This training enables LLMs to predict subsequent words in a sentence, generate human-like text, and even engage in conversations. For instance, models like OpenAI's GPT series have demonstrated remarkable proficiency in generating text that is often indistinguishable from that written by humans [30].

State-of-the-Art Developments

The field of LLMs is rapidly evolving, with significant advancements enhancing their capabilities and efficiency. A notable development is the introduction of models that achieve high performance with reduced computational resources. For example, DeepSeek's R1 model employs a "mixture of experts" approach, dynamically activating relevant networks based on tasks and feedback. This method allows the model to adjust over 671 billion parameters intermittently, significantly reducing computational requirements without compromising performance. Remarkably, the R1 model has excelled in areas such as mathematics and coding, ranking highly on platforms like Chatbot Arena [31].

Another significant trend is the open-source movement within the AI community. DeepSeek's R1 model, for instance, is open source, enabling developers worldwide to modify and enhance its capabilities. This approach fosters collaboration and accelerates innovation, as seen with other models like Meta's Code Llama, which is designed specifically for code generation and understanding [32].

Furthermore, the global landscape of AI development is becoming more diverse. Chinese companies are making significant strides in AI technology. Firms such as ByteDance, Alibaba, Baidu, and Tencent have developed competitive LLMs like Doubao, Qwen, Ernie Bot, and Hunyuan, respectively. These models are not only technologically advanced but also cost-effective, making AI applications more accessible. The adoption of open-source strategies and innovative frameworks, such as the "mixture of experts," has been pivotal in these developments [33].

Implications and Future Directions

The advancements in LLMs have profound implications across various sectors. The reduction in computational costs and the open-source nature of modern models democratize AI technology, allowing a broader range of organizations to develop and deploy AI solutions. This democratization is poised to accelerate innovation in fields like healthcare, finance, and education, where AI can offer personalized and efficient services.

However, these developments also raise important considerations regarding ethics, security, and the global balance of technological power. As AI models become more sophisticated and accessible, ensuring their responsible use becomes paramount. This includes addressing potential biases in AI outputs, safeguarding against malicious applications, and fostering international collaboration to set standards and regulations.

APPENDIX 2: GRAPH ONTOLOGY EXTENDED EXAMPLE

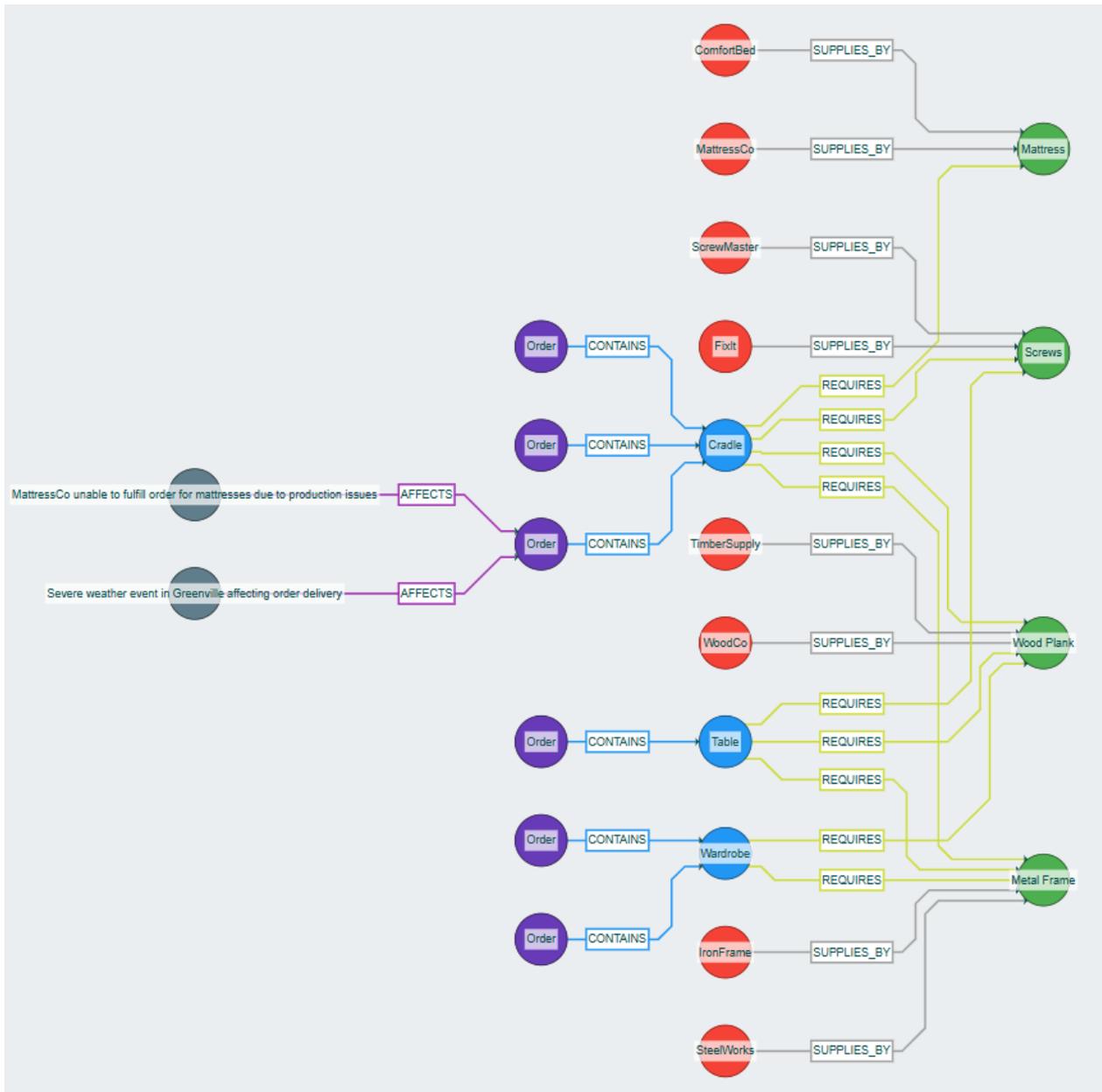


FIGURE 8: EXAMPLE OF ONTOLOGY

6. REFERENCES

- [1] Wooldridge, M. (2020). *An Introduction to MultiAgent Systems* (2nd ed.). Wiley.
- [2] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- [3] Shoham, Y., & Leyton-Brown, K. (2009). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press.
- [4] Foerster, J. N., et al. (2018). *Deep multi-agent reinforcement learning*. arXiv:1810.05587.
- [5] Jiang, J., et al. (2020). *Graph convolutional reinforcement learning for multi-agent cooperation*. Advances in Neural Information Processing Systems.
- [6] Bommasani, R., et al. (2021). On the Opportunities and Risks of Foundation Models. arXiv:2108.07258.
- [7] Yang, R., et al. (2022). Game-theoretic and auction-based approaches for resource allocation in multi-agent systems. *ACM Computing Surveys*.
- [8] Li, T., et al. (2021). Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine*.
- [9] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
- [10] Chen, H. (2024). *Understand the LLM Agent Orchestration*. Medium. Retrieved from <https://medium.com/scisharp/understand-the-llm-agent-orchestration-043ebfaead1f>.
- [11] Chase, H. (2023). *LangChain: Building applications with LLMs*. Available at: <https://python.langchain.com>
- [12] OpenAI. (2023). *OpenAI API Documentation*. Available at: <https://platform.openai.com/docs/>
- [13] Johnson, J., Douze, M., & Jégou, H. (2017). *Billion-scale similarity search with GPUs*. *IEEE Transactions on Big Data*.
- [14] Medium. (2020). *Importing RDFs/OWLs Ontologies into Neo4j*. Retrieved from <https://medium.com/neo4j/importing-rdfs-owl-ontologies-into-neo4j-23e4e28ebbad>
- [15] Lemaignan, S., & Siadat, M. (2021). *MASON: A Proposal For An Ontology Of Manufacturing Systems*. Retrieved from <https://www.semanticscholar.org/paper/MASON%3A-A-Proposal-For-An-Ontology-Of-Manufacturing-Lemaignan-Siadat/0e7b6ba7763ed72666bc752112324dd2e73b91d6>
- [16] Sapel, P., & Molinas Comet, L. (2024). *A review and classification of manufacturing ontologies*. Retrieved from <https://www.semanticscholar.org/paper/A-review-and-classification-of-manufacturing-Sapel-Comet/033e3e544e484d9db6e87535e685d88695319172>
- [17] Fraga, A., & Vegetti, M. (2024). *Ontology-based solutions for interoperability among product*. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0278612524002607>
- [18] Indicium Tech. (2023, July 10). Building agents that can call APIs with the OpenAPI Specification. Medium. Retrieved from <https://medium.com/indiciumtech/building-agents-that-can-call-apis-with-the-openapi-specification-c45b991691af>
- [19] Indicium Tech. (2023, July 10). Building agents that can call APIs with the OpenAPI Specification. Medium. Retrieved from <https://medium.com/indiciumtech/building-agents-that-can-call-apis-with-the-openapi-specification-c45b991691af>
- [20] GetAmbassador.io. (2022, October 5). OpenAPI Specification structure & best practices. GetAmbassador.io. Retrieved from <https://www.getambassador.io/blog/openapi-specification-structure-best-practices>
- [21] LangChain Documentation. Source: <https://python.langchain.com/docs/integrations/tools/openapi/>

- [22] Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.
- [23] Brown, T., et al. (2020). Language Models are Few-Shot Learners. NeurIPS.
- [24] Radford, A., et al. (2019). GPT-2: Better Language Models and Their Implications. OpenAI.
- [25] Vaswani, A., et al. (2017). Attention is All You Need. NeurIPS.
- [26] Liu, P., et al. (2021). Pre-train Prompt Tune: Exploring the Best Way to Adapt Pre-trained Language Models. ACL.
- [27] Bommasani, R., et al. (2021). On the Opportunities and Risks of Foundation Models. Stanford.
- [28] Wei, J., et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. NeurIPS.
- [29] LangChain Documentation. (2023). Building LLM-Powered Applications. LangChain.
- [30] Wikipedia. (n.d.). Large Language Model. Retrieved from https://en.wikipedia.org/wiki/Large_language_model
- [31] DeepSeek AI. (2024). DeepSeek’s R1 Model and the Mixture of Experts Approach. The Wall Street Journal. Retrieved from <https://www.wsj.com/tech/ai/deepseek-ai-how-it-works-725cb464>
- [32] Meta AI. (2023). Code Llama: Advancing Large Language Models for Code Generation. Meta AI Blog. Retrieved from <https://ai.meta.com/blog/code-llama-large-language-model-coding>
- [33] Reuters. (2025). *China’s AI Firms Take Spotlight with Deals, Low-Cost Models*. Reuters. Retrieved from <https://www.reuters.com/technology/artificial-intelligence/chinas-ai-firms-take-spotlight-with-deals-low-cost-models-2025-02-14>