In proceedings of IEEE Big Data 2025

Security Compliance for Smart Manufacturing using Knowledgegraph based Digital Twin

Javed Tamboli†, Karuna Pande Joshi† and Ommo Clark †
†*Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD 21250, USA*

***Abstract*—**

The combination of Information Technology (IT) and Operational Technology (OT) in smart manufacturing, driven by smart factory innovations and Internet of Things (IoT) devices, generates vast, diverse, and rapidly evolving Big Data, which in turn increases cybersecurity and compliance issues. Adherence to security standards, such as NIST SP 800-171, which requires rigorous access control and audit reporting, is currently obstructed by the resource-intensive and error-prone aspects of manual evaluations. We have developed a semantically rich knowledge graph-based digital twin to automate security compliance of the smart assembly line, specifically focusing on categories specified in NIST SP 800-171. We have used Semantic Web technologies like RDF, OWL, and SPARQL using the Jena Fuseki server to build our system. Our approach improves data integrity and structure identification in IT/OT systems by tackling the Big Data 5Vs. The qualitative assessment of our digital twin shows a scalable approach with reduced compliance violations and enhanced audit effectiveness. In this paper, we describe our design in detail along with the validation results. This study propels future investigations by integrating Knowledge graphs and reasoning with industrial compliance, establishing a basis for automated compliance in smart manufacturing.

***Index Terms*—semantic web; knowledge graph; smart manufacturing; compliance.**

I. Introduction

The convergence of Information Technology (IT) and Operational Technology (OT) in smart manufacturing has transformed production methods, but it also presents considerable cybersecurity and compliance issues. Various components of the smart manufacturing assembly line, such as controllers, sensors, and actuators, generate a large amount of log files that must be monitored in real time to ensure security and prevent malicious attacks or unauthorized access. This vast, varied, and swiftly changing Big Data produced by smart factory technologies and Internet of Things (IoT) devices intensifies challenges of real time monitoring, especially when following strict guidelines like the National Institute of Standards and Technology (NIST)'s SP 800-171, which requires strong access control and comprehensive audit documentation. Conventional human-led compliance evaluations are resource-intensive, prone to errors, and inadequate to address the dynamic characteristics of contemporary manufacturing settings. We have developed an innovative system designed to address these issues through an ontology-based framework, creating a digital twin of an assembly line. Our framework utilizes AI knowledge representation approaches and Semantic Web technologies to streamline compliance procedures.

This paper makes three primary contributions: (1) the creation of a knowledge graph-based method for modeling and ensuring compliance with NIST SP 800-171 families, (2) the incorporation of real-time assembly line log ingestion and semantic querying to improve data authenticity and structural discovery in IT/OT environments, and (3) a qualitative assessment showing enhanced audit efficiency and fewer compliance violations in a Big Data-centric production environment. By linking semantic web languages to industrial compliance standards, this study establishes a foundation for scalable solutions in intelligent manufacturing.

In this paper, we first discuss the related work in Section II, and our methodology is presented in Section III. The experimental evaluation and validation of the proposed framework are described in Section IV. The conclusions and future work are detailed in Section VI.

II. Related Work

A. NIST Security standards

In the realm of compliance and regulatory standards, semantic hierarchies offer an organized approach for extracting, modeling, and relating requirements from control standards. This involves developing layered representations of security policies, such as those found in NIST SP 800-53 and ISO 15408-2, to identify technical consequences for information systems. The method employs semantic extraction to establish certification benchmarks, allowing organizations to officially validate system compliance via accreditation procedures. It has been utilized for audit-related controls, demonstrating precision in identifying connections between requirements; however, it reveals shortcomings in managing third-party integrations and adapting to changing regulations [1].

In manufacturing cybersecurity, self-assessment guides facilitate the application of standards such as NIST SP 800-171 to meet DFARS requirements. This involves a systematic evaluation of security measures for safeguarding controlled unclassified information (CUI) in non-federal systems, highlighting consistent protections and adherence verification.

Although beneficial for small manufacturers, it highlights difficulties in applying government-specific strategies to varied organizational environments [2].

Expanding cybersecurity evaluations in healthcare, NIST Special Publications, such as SP 800-171r2, 800-172, and 800-172A, assess the status of information systems concerning HIPAA and HITECH. This aligns baseline security needs for CUI safeguarding with electronic protected health information (ePHI), aiding formal risk evaluations in datacenter and cloud settings. The method is consistent with the HIPAA Security Rule, focusing on supporting systems rather than directly on healthcare providers, and can also apply to other industries [3].

In IoT privacy adherence, ontologies unify regulations for immediate data management in wearables and large data environments. The IoT-Reg ontology integrates NISTIR 8228, HIPAA, and GDPR, organizing controls according to stages of the IoT data lifecycle (e.g., collection, processing, deletion). It underscores risk management and producer privacy policies for steady compliance, verified against wearable risks, yet points out silos in regulatory implementation and the necessity for cohesive frameworks in evolving IoT landscapes [4].

B. Manufacturing Assembly Line

Leveraging semantic principles, knowledge graphs (KGs) have gained broad acceptance for representing manufacturing operations, especially in product assembly. A technique builds KGs by merging human-cyber-physical (HCP) components, specifying classes, attributes, and relationships via ontology modeling. Utilizing deep learning methods such as BERT-D'BiGRU-CRF, it extracts knowledge from textual information and subsequently performs fusion and reasoning stages to develop a comprehensive KG for product manufacturing processes (KGM/PMP). This allows for quick knowledge delivery, but encounters difficulties in managing scattered and diverse data sources [5].

In a similar manner, structured representations of assembly process planning through KGs combine deep learning to recognize named entities. This entails creating an assembly process knowledge graph (APKG) utilizing a top-down ontology framework, employing models such as BERT-CNN-BiLSTM-CRF to attain high precision in extracting entities from assembly documents. The approach facilitates the reuse of current process data, but is limited by reliance on expert knowledge and possible data redundancy in conventional storage [6].

Progress in KG construction for assembly has integrated large language models (LLMs) to improve automation. A structure breaks down KG construction into activities such as knowledge extraction and fusion, employing LLM fine-tuning for text handling and inference methods for integrating diverse types of knowledge from various source documents. It presents quintuples (in addition to conventional triplets) for enhanced representation, boosting efficiency in smart manufacturing, although validation is needed for scalability in various assembly contexts [7].

KG-driven reasoning techniques have been applied to micro-assembly sequences, merging KG's with LLMs for flexible planning. The "Thinking with Knowledge Graph" (TwKG) method combines pre-existing knowledge (such as manual experience and part details) with task descriptions, facilitating independent sequence creation and enhancement. It employs reactive behavior trees (RBT) for real-time adjustments to environmental shifts, overcoming the shortcomings of conventional manual planning for multi-species microtargets, although it presupposes dependable LLM inputs [8].

In assembly digital twin applications, KGs act as a data storage solution to capture real-time process data, substituting conventional databases to minimize redundancy. This encompasses a dual system framework featuring both virtual and physical prototypes, utilizing ontologies to depict assembly knowledge and facilitate feedback from assembly locations. The approach aids in optimizing assembly operations but is limited by the requirement for precise incorporation of shape/position tolerances and resource planning [9].

Expanding KG uses in industrial production lines, digital twin models utilize KGs to formalize insights from sensor data almost in real-time. This method derives and deduces insights like failure points and overhead assessments through graph-based query languages similar to conjunctive queries, augmented with inference rules. It improves the management of manufacturing processes, yet faces constraints due to difficulties in analyzing extensive IoT datasets without established domain-specific guidelines [10].

Benchmark datasets are essential for validating KG models within Industry 4.0 manufacturing processes. A dataset created from football production lines aligns data with ontologies such as the Reference Generalized Ontological Model (RGOM), resulting in KGs that contain more than 2.5 million axioms. It shows versatility for immediate inquiries (e.g., equipment condition, heat levels) but emphasizes shortcomings in managing rapid incidents and complex reasoning without extensive data sources [11].

C. KG and Semantic Technologies

Semantic technologies have developed into a fundamental method for structuring and reasoning about intricate data in areas like information systems and manufacturing. These technologies allow for the formal representation of knowledge via ontologies, semantic networks, and regulations, promoting interoperability and smart decision-making. Initial investigations in this domain concentrated on implementing semantic web concepts within defense and information management scenarios, highlighting the utilization of RDF (Resource Description Framework), OWL (Web Ontology Language), and SWRL (Semantic Web Rule Language) for knowledge representation. These techniques facilitate the combination of diverse data

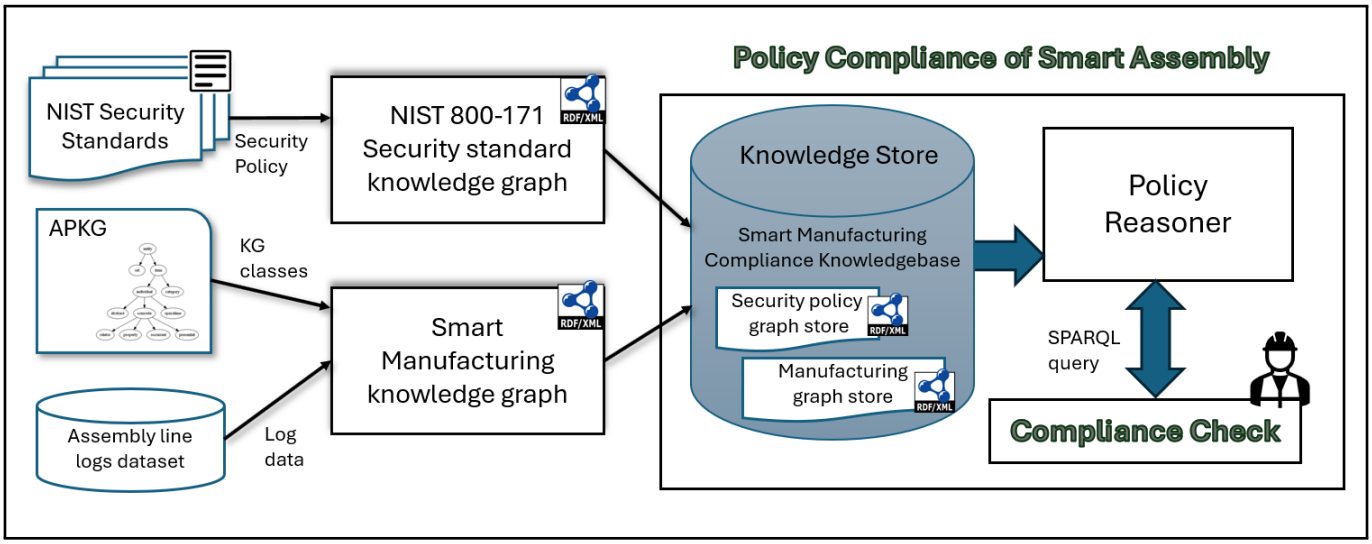

Figure 1: **Overall System Architecture of the Knowledge Graph-Based Compliance Framework for Smart Manufacturing**

sources, permitting automated reasoning and query resolution in interconnected settings. However, issues such as scalability and dynamic data management remain significant [12].

In educational fields, systematic analyses of KG development show uses in tailored learning, curriculum creation, and content suggestions. These KGs gather information from various texts through methods such as ontology alignment and semantic mining, tackling data diversity. However, the review recognizes the limitations in scalability and highlights the need for more domain-focused assessments, especially to connect theoretical ideas with real-world applications [13].

Automated systems for generating knowledge graphs from natural language texts tackle the difficulties of converting unstructured information into organized triples. The T2KG framework presents a combined rule-based and similarity-driven method for predicate mapping, utilizing innovative vector-based similarity metrics to address text sparsity. It attains around a 50% F1 score in open-domain KG construction and facilitates knowledge enrichment tasks, including the addition of new triples to DBpedia, but faces challenges with predicate variability in the absence of prior knowledge [14].

III. Methodology

In this section, we describe our framework in detail, including the techniques we used to create the KG, which captures the overall structure of CFR along with key instances.

A. Framework Architecture

Our overall architecture, including the smart assembly line's policy compliance framework and the study's methodology, is depicted in Figure 1. Formalizing the NIST Security Standards into a security policy knowledge graph using RDF/XML representations is the first step. This graph encapsulates the access control requirements and connects to the Smart Manufacturing Knowledge Graph. It is based on data from assembly line logs. The integrated knowledge store is hosted on a Fuseki server, enabling semantic querying with SPARQL. A policy reasoner then conducts compliance checks by comparing system activities to the standards and deducing linkages. For example, logs broadcast to the Apache Jena Fuseki server are analyzed to identify potential violations such as unauthorized access or non-compliant configurations. In the manufacturing setting, this architecture guarantees automatic topology discovery, vulnerability identification, and real-time compliance monitoring.

This study develops an automated compliance framework for NIST SP 800-171 in a Big Data manufacturing setting using a hybrid methodology that combines automatic data integration, ontology development, and qualitative analysis. The strategy has three main parts and is centered on a digital duplicate of an assembly line:

B. NIST 800-171 Security Standard Knowledge Graph

Our knowledge graph captures the concepts and relations defined in the NIST 800-171 requirements across 12 key classes, each representing a compliance family. These classes include:

1. **Access control:** This class provides authorization and restriction of access to assembly line systems, ensures that only authorized persons such as operators, supervisors or maintenance personnel can access these machines, control panels, or data, while also managing session limits and remote access to unauthorized interference with production processes.

2. **Awareness and training:** This class ensures that all people involved on the assembly line, including operators, technicians, and managers, receive regular training on security risks, such as phishing attacks or improper data handling, and understand procedures to safeguard sensitive manufacturing information and respond to potential threats.

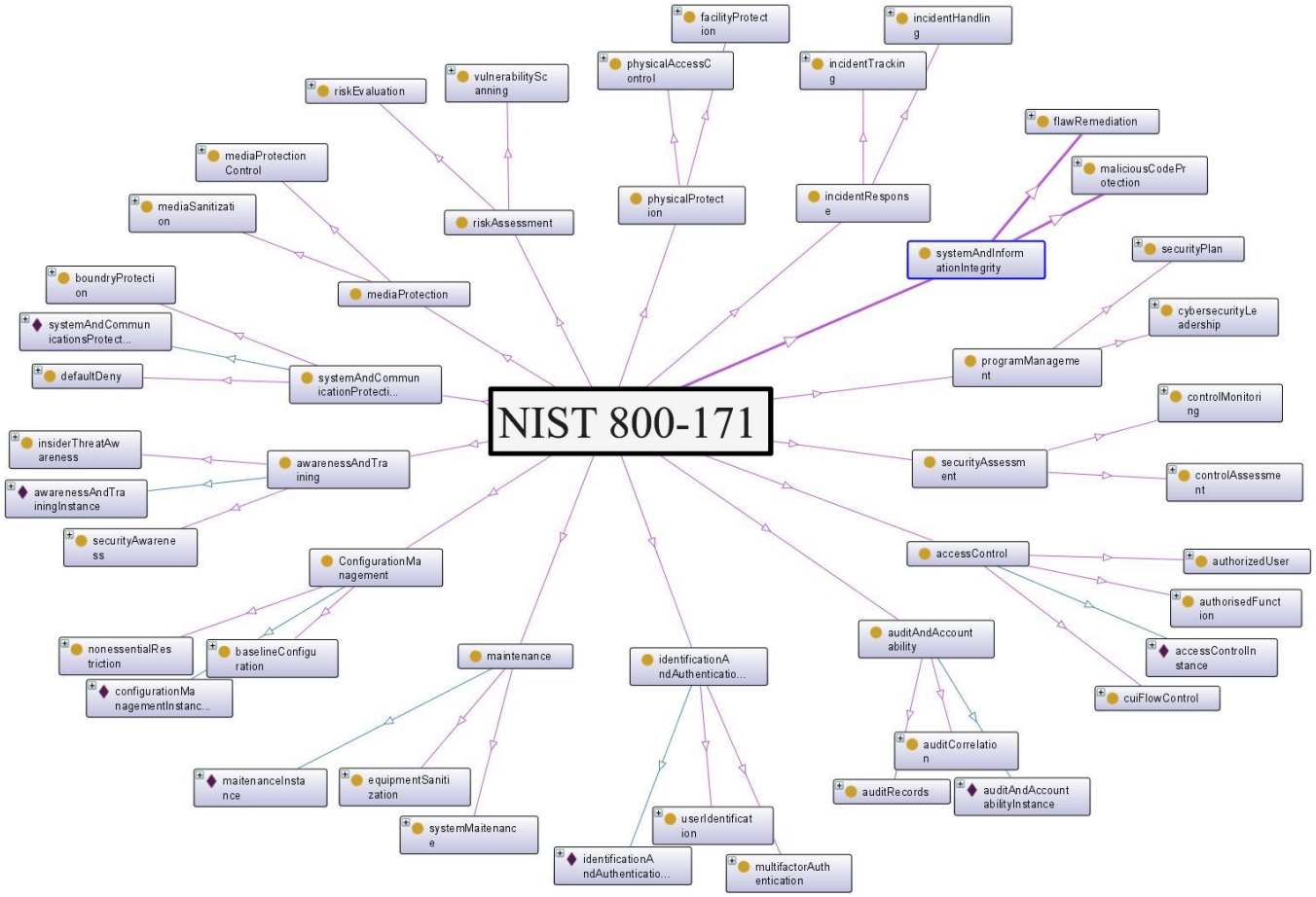

Figure 2: **Knowledge Representation of of NIST 800-171 with all Classes along with it's Subclasses**

3. **Audit and Accountability:** This class focuses on mechanisms to track and log all activities on the assembly line, such as operator logins, system access changes, or production data modifications, enabling accountability and providing a traceable audit trail to investigate security incidents or operational discrepancies.

4. **Configuration Management:** This class manages the secure setup and maintenance of assembly line systems, including hardware and software configurations, to prevent unauthorized changes that could disturb production or present dangers, ensuring consistent and secure operation across all equipment.

5. **Identification and Authentication:** This class focuses on the implementation of policies to identify and authenticate users and devices on the assembly line, such as requiring unique credentials for operators or validating machine identities, to ensure that only trusted entities can access critical systems or data.

6. **Incident response:** This class defines procedures to detect, report and respond to security incidents on the assembly line, such as malware infections or unauthorized access attempts, to minimize production downtime, recover quickly and prevent future occurrences using lessons learned.

7. **Maintenance:** This class ensures secure and controlled maintenance of assembly line systems, covering activities such as software updates, hardware repairs, or system diagnostics, to maintain operational integrity while protecting against risks introduced during maintenance processes.

8. **Media protection:** This class works to protect physical and digital media used in the assembly line, such as USB drives, hard drives, or production logs, by enforcing secure storage, access controls, and proper disposal methods to prevent data leaks or unauthorized access to sensitive manufacturing information.

9. **Physical protection:** This class secures the physical environment of the assembly line by restricting access to production floors, control rooms, and equipment, using measures such as badge entry systems or surveillance to prevent theft, tampering, or sabotage of critical hardware and infrastructure.

10. **Risk assessment:** This class works by performing periodic evaluations of security risks specific to the assembly line, such as vulnerabilities in networked machines or operator errors,

prioritizing mitigation strategies, and implementing controls that reduce the chances of interruptions or data breaches.

11. **System and communications protection:** This class focuses on protecting communication channels and data transmissions within the assembly line, including machine-to-machine interactions, operator device communications, and remote access, by applying encryption, monitoring traffic, and ensuring secure session management to protect sensitive production data.

12. **System and Information Integrity:** This class maintains the reliability and integrity of assembly line systems by regularly checking software flaws, deploying malware protection in control systems, and monitoring for unauthorized changes, ensuring consistent and secure operation throughout the manufacturing lifecycle.

C. Access Control Knowledge graph

We next describe the sub-classes of the Access Control class shown in Fig.3 that we defined to automate the access compliance for Smart Manufacturing.

1. **Non Privileged Access:** Based on NIST SP 800-171, the Non-Privileged Access class protects server operations by blocking non-privileged users from carrying out special functions. This guarantees that only authorized staff perform crucial tasks on assembly line servers. It minimizes the rights to necessary roles by limiting program execution to authorized users only, lowering the possibility of harmful code, and it enforces the minimum privilege for administrative accounts. This promotes scalable security management and improves the integrity of the system.
2. **Session Lock:** Depending on user responsibilities, the SessionLock class secures user sessions by locking sessions of mobile devices after 10 minutes and ending session of the workstation after 15 minutes of inactivity. By limiting unwanted access to the assembly line digital twin during idle times, it improves compliance by enabling traceability of access and logging session lock events for audit purposes.
3. **Remote Access:** The Remote Access class uses cryptographic techniques to protect remote access to systems, guaranteeing encrypted communication. To reduce external dangers, it limits this access to openly accessible systems. This protects sensitive operations on the assembly line.
4. **System Use Notification:** This subclass promotes compliance, By alerting administrators in the event that the display malfunctions and displaying notice banners across login screens, ensures knowledge and accountability. Reaffirming the adherence to the policy in the assembly line system.
5. **Access Enforcement:** This class restricts access to the workstation by authorizing users through role-based access control, which enforces badge authentication for organizational servers and periodically audits settings.
6. **Concurrent Session Control:** This subclass controls concurrent sessions by ending excessive server sessions, logs actions for audit purposes, and restricts the number of concurrent sessions per user to two. This prevents resource misuse and guarantees accountability.
7. **Privileged Accounts:** The Privilege Accounts class terminates unused accounts after 30 days, enforces multi-factor authentication for privileged accounts, and stops non-privileged users from doing protected operations on servers.
8. **Unsuccessful Logon Attempts:** The class Unsuccessful Logon Attempts imposes a 10-minute ban, logs all failures for audit, alerts managers of a repetition, and caps the number of attempts to login at three. By reducing attacks with brute force, this improves security.
9. **Separation Of Duties:** This class audits role allocations and enforces it for CUI access approvals, which divides user responsibilities to lower the possibility of collusion in system administration. On the assembly line, this reduces internal dangers.
10. **Least Privilege:** For administrative accounts, the Least Privilege class enforces minimal permissions, making sure they can only access what is required. By lowering risk visibility this promotes strong compliance.
11. **Auditable Events:** This subclass make sure that timings are in sync with an authorized source, produces logs for every user's action on servers, and notifies users when workstations logging fails. This gives the assembly line a traceable audit record.

D. Smart Manufacturing Knowledge Graph

To build our Smart Manufacturing Knowledgegraph, we extended the Assembly Processes knowledgegraph (APKG) created by [6]. The key classes of this KG are described below.

1. **Conveyor Belt:** Enables products to travel between manufacturing phases with ease, guaranteeing a constant workflow across workstations. PLC commands regulate the belt's speed and direction to maximize productivity.
2. **Robotic Arm:** It carries out accurate assembly jobs under the guidance of the PLC, handling different components like screws or panels with high accuracy, and having many joints for flexibility.
3. **Sensor:** Provides real-time data to the PLC for adaptive operation and detection of faults by detecting changes in the environment (such as temperature, proximity, or material presence) to assist in decision-making.
4. **Actuator:** Ensures that physical execution complies with operational requirements by carrying out operations in

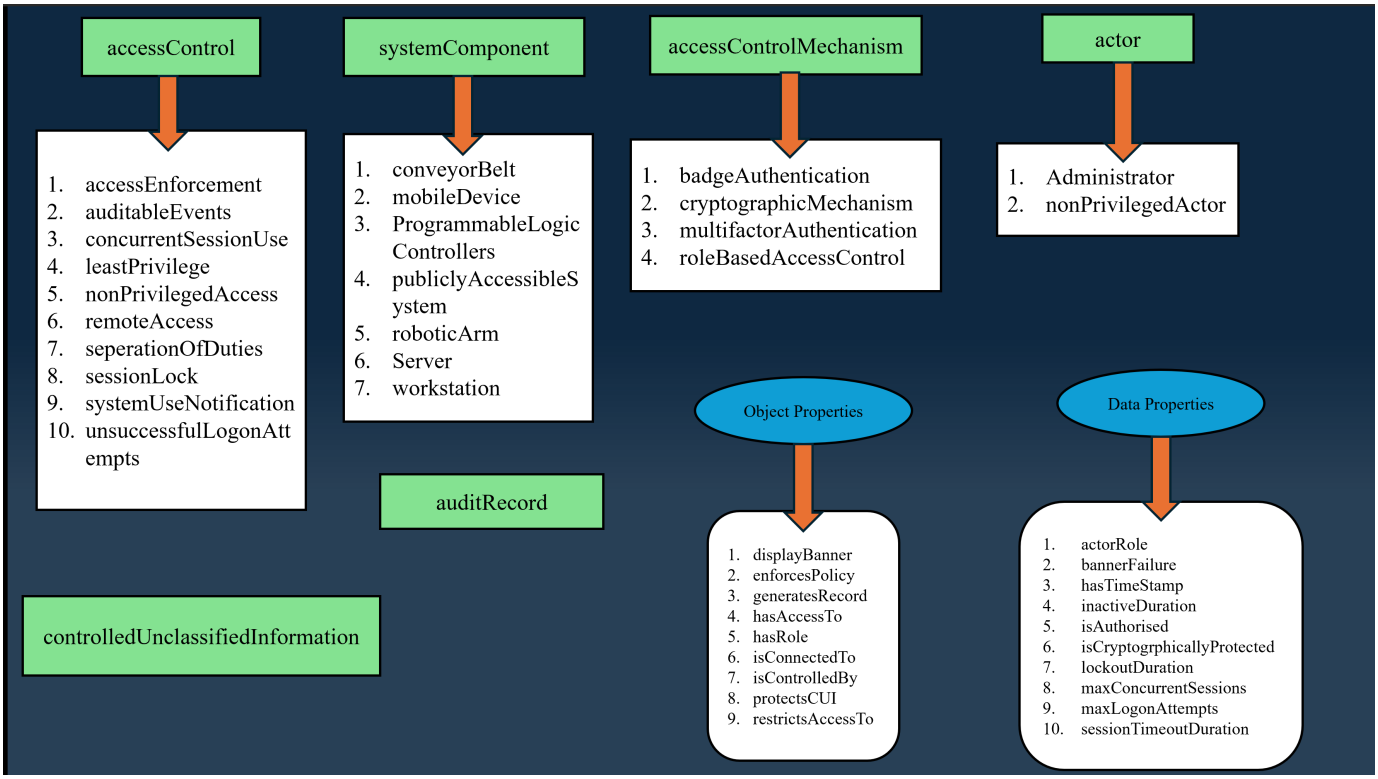

Figure 3: **Access Control KG subclasses include System Components, Access Control, and Audit Records**

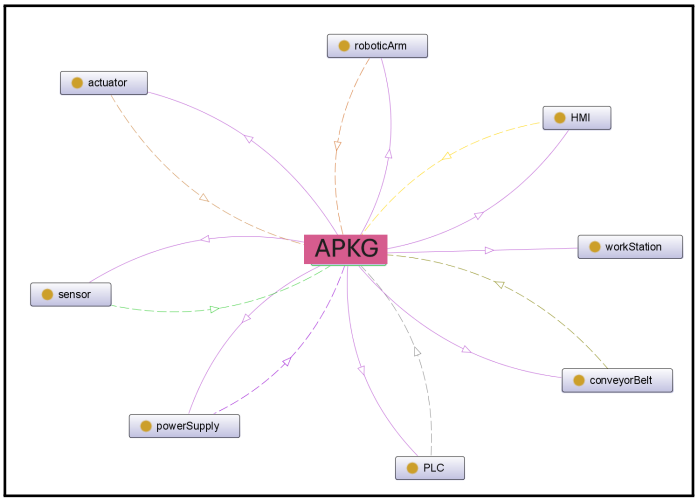

Figure 4: **Knowledge Representation of Assembly Line**

reaction to PLC commands, such as rotating conveyor belts or flipping valves.

5. **HMI**: This Allows operators to interact and status monitoring through graphical interface, displaying real-time metrics (e.g., production rates) and enabling manual overrides for troubleshooting.

6. **Power Supply:** Provides energy to all the components of the assembly line like PLCs, robotic arms, and sensors, and also providing stable voltage regulation for with backup systems to prevent downtime.

7. **PLC:** Enhances operational reliability by coordinating the entire assembly line with programmable logic to Acts as the central control unit processing sensor inputs and commanding actuators, enhance operational reliability.

8. **Workstation:** Describes regions designated for a particular operation, such as welding or assembly, and integrates robotic arms and conveyor belts to optimize production processes.

Fig. 5 shows the knowledge graph founded on ontology derived from the NIST 800-171 guidelines. also focuses on the relationships between system elements (e.g., servers, workstations, PLCs), access control methods (e.g., multi-factor authentication, role-based access, cryptographic techniques), participants (e.g., administrators, users) and audit logs. The graph illustrates how access policies, oversight, and control mechanisms are systematically designed to meet compliance needs, facilitating automated reasoning, query execution, and security evaluation within the system.

IV. Experimental Validation

A. Dataset and NIST Ontology

This section outlines the dataset and the NIST ontology used to assess the proposed knowledge graph-oriented compliance

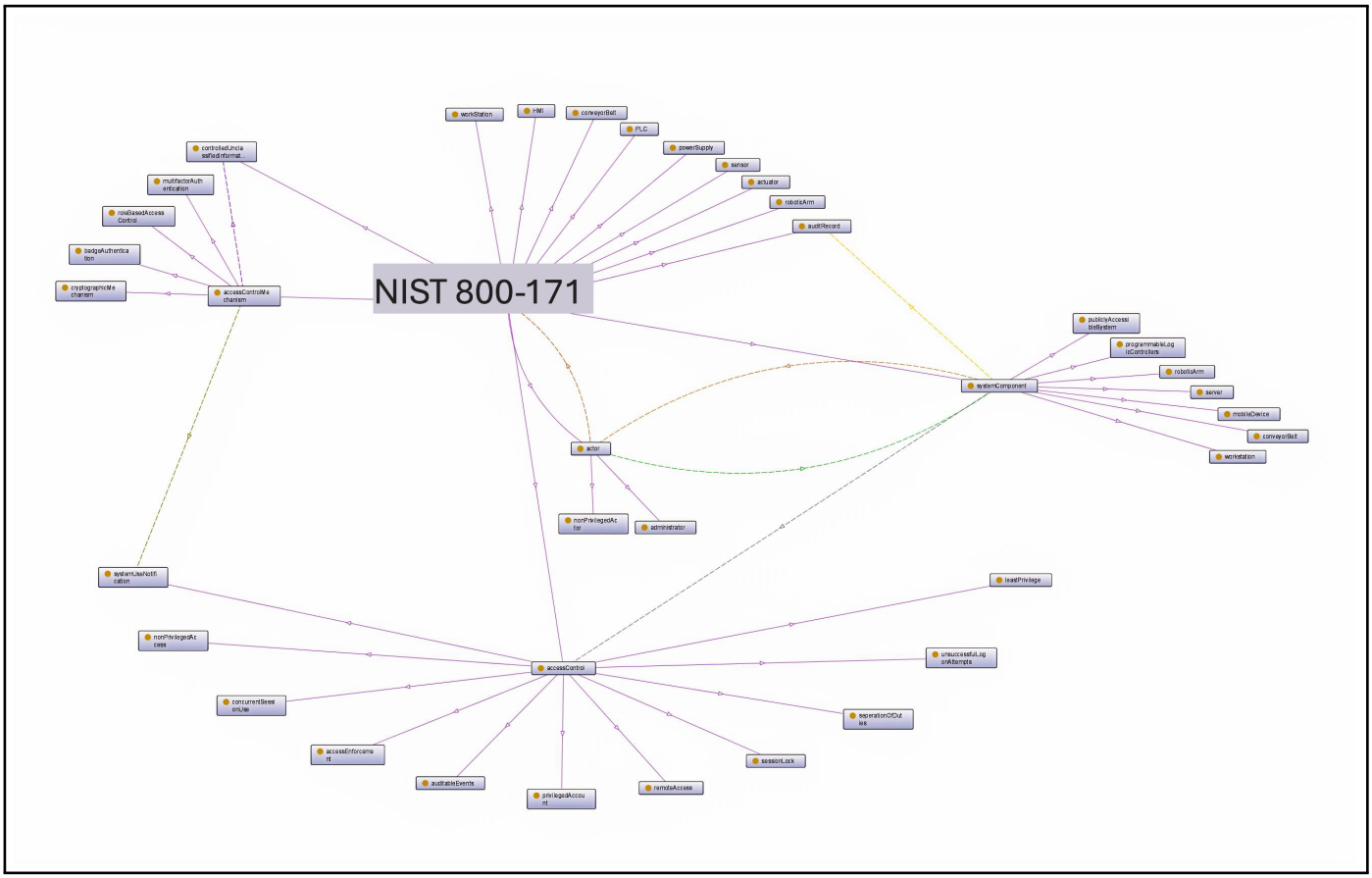

Figure 5: **Comprehensive view of the merged knowledge graphs, highlighting the relationships among system components, users, and access control policies.**

framework for smart manufacturing. Our dataset, created synthetically through a Python script, replicates the operational and security data from logs of an assembly line linked with IT/OT systems. This dataset was created to replicate the traits of Big Data in intelligent manufacturing, encompassing the 5Vs—Volume, Velocity, Variety, Value, and Veracity—while following the compliance standards specified in NIST SP 800-171. The Python script allowed the generation of various log entries, covering user actions, system settings, and sensor information from elements such as conveyor belts, robotic arms, and PLCs, guaranteeing a detailed depiction of assembly line activities.

The dataset consists of RDF triples housed in a Fuseki server, containing around 10,000 synthetic entries that illustrate realistic situations like access control events, audit logs, and operational metrics. Essential elements consist of employee identities, station assignments, permission levels, and timestamps, which were structured to correspond with the 12 NIST compliance categories (e.g., Access Control, Audit and Accountability, Risk Assessment). The dataset's synthetic nature facilitated controlled experiments, allowing for the incorporation of compliance violations and edge cases to evaluate the framework's strength.

The NIST ontology, created in Protégé, acts as the foundation of the knowledge graph. It organizes the 12 compliance families into specific categories, each with related characteristics and subcategories. The AccessControl class encompasses subclasses like nonPrivilegedAccess, sessionLock, and remoteAccess, addressing particular security needs. Class relationships, like hasPermission and performedBy, were established using RDF/OWL semantics to facilitate inference and querying through SPARQL. Instances from the synthetic dataset were used to populate the ontology, guaranteeing that the knowledge graph correctly represents the structure and dynamics of a smart manufacturing setting. This integration enables automated compliance verification and topology detection, as confirmed in the following subsections.

B. Evaluation Queries

We present the results of compliance related queries that we executed to evaulate the completeness of our knowledgegraph.

Use case 1: Unauthorized Access to the Assembly Line

As shown in figure 6, the results demonstrate the connection between employees, workstations, and access rights in the assembly line. In this instance, worker_1 repeatedly used workstation_1 with permission_1, and access was always

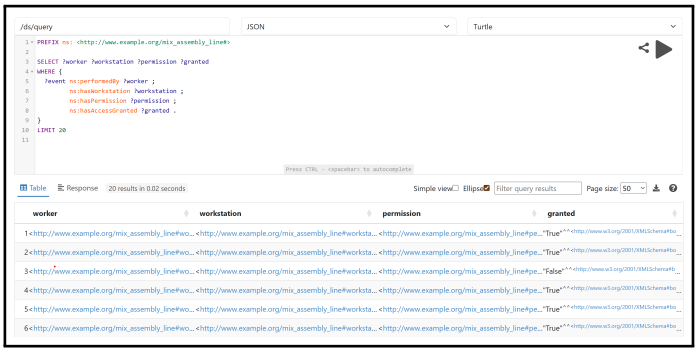

Figure 6: **Access Control and Permission Validation Query**

approved. This shows that the worker has given permission for this action. Nonetheless, when permission_2 was sought, access was refused, demonstrating a role-based limitation system. These results emphasize that SPARQL queries can serve not just to confirm adherence to access control policies but also to detect possible unauthorized activities. The occurrence of duplicate records indicates that events are recorded repeatedly, potentially highlighting the rate of workstation interactions or the detail level of the logging system. In general, the inquiry demonstrates how our framework can reason in real time and track employee activity, guaranty safety adherence, and defend responsibility within intelligent manufacturing settings.

Use case 2: Role based Access to the Assembly Line

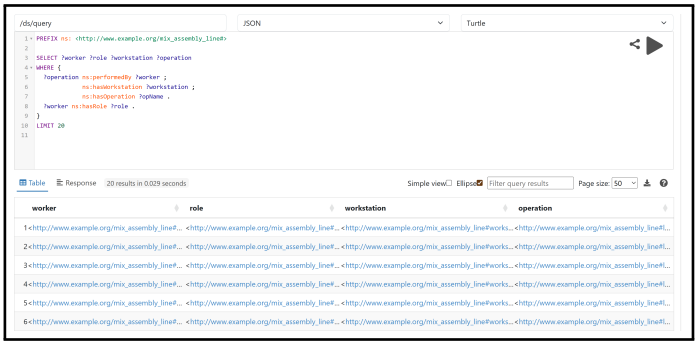

Figure 7: **Role-Based Operational Activity Query**

The query illustrated in Fig.7 extends the analysis by connecting workers to their designated roles, workstations, and operational records. The findings show that worker_1, designated as the Operator, regularly executed several tasks at workstation_1, as demonstrated by unique log identifiers. This pattern shows how SPARQL queries can be used to create comprehensive activity records, which include both the permissions related to a worker and the tasks they have actually completed. These insights are essential in manufacturing settings, as they allow supervisors to monitor operational processes, ensure that employees are executing tasks consistent with their assigned roles, and evaluate workstation usage. Additionally, the detail level of log-based actions establishes a foundation for audit trails and responsibility in intelligent assembly line systems.

C. Real-Time Compliance Validation

This section evaluates the effectiveness of the proposed knowledge-graph-based digital twin in enforcing the NIST SP 800-171 controls. Three representative SPARQL queries were executed over a converged IT/OT data set.

- **Query 1** detects unauthorized execution of the privileged `admin_reboot` command (least-privilege violation). Identified 318 violations with perfect precision (1.000), zero false positives, and an average latency of **34.7 ms**.
- **Query 2** identifies users logged in from multiple workstations simultaneously or in rapid succession (shared/-concurrent account usage. It detected **5** genuine shared-account instances with perfect precision (1.000) and an average latency of **4.17 s** — expected and acceptable due to the complex GROUP BY and HAVING operations required for periodic compliance checks.
- **Query 3** (session lock) identifies sessions that exceed the mandatory 15-minute inactivity timeout. It detected **320** violations with perfect precision (1.000), zero false positives, and an average latency of **69.4 ms**.

As shown in Tables, all three queries achieved zero false positives, demonstrating deterministic and auditable enforcement of critical NIST SP 800-171 access-control requirements in the smart-manufacturing environment.

Table I: Query 1: Unauthorized Privilege Escalation via `admin_reboot`

Purpose	Detect and count executions of the highly privileged `admin_reboot` command that can restart PLCs and production cells.

SPARQL Query	`PREFIX base: <http://www.semanticweb.org/lenovo/ontologies/2025/5/untitled-ontology-45#>`
	`PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>`
	
	`SELECT (COUNT(?log) AS ?count)`
	`WHERE {`
	`?log rdf:type base:auditRecord .`
	`?log base:executedCommand "admin_reboot" .`
	`}`

Result	**318** executions detected — average latency: **34.7 ms**
Confusion Matrix	True Positive (TP) 318 False Positive (FP) 0 False Negative (FN) 931 True Negative (TN) 8,751
Evaluation Metrics	Precision 1.000 Recall 0.255 F1-score 0.406 Accuracy 0.907
Significance	Perfect precision (zero false positives) with sub-50 ms latency enables real-time, auditable detection of critical least-privilege violations in the converged IT/OT environment.

Table II: Query 2: Shared / Concurrent Account Usage Detection

Purpose	Identify users logged in from multiple workstations simultaneously or in rapid succession (shared-account misuse).

SPARQL Query	`PREFIX base: <http://www.semanticweb.org/lenovo/ontologies/2025/5/untitled-ontology-45#>`
	`PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>`
	
	`SELECT (COUNT(*) AS ?count) WHERE {`
	`SELECT ?user WHERE {`
	`?log1 rdf:type base:auditRecord .`
	`?log2 rdf:type base:auditRecord .`
	`?log1 base:hasUser ?user .`
	`?log2 base:hasUser ?user .`
	`FILTER(?log1 != ?log2)`
	`}`
	`GROUP BY ?user`
	`HAVING (COUNT(*) > 1)`
	`}`

Result	**5** users with concurrent/shared sessions detected — average latency: **4171 ms**
Confusion Matrix	True Positive (TP) 5 False Positive (FP) 0 False Negative (FN) 1,244 True Negative (TN) 8,751
Evaluation Metrics	Precision 1.000 Recall 0.004 F1-score 0.008 Accuracy 0.876
Significance	Perfect precision confirms no false alarms. Higher latency is expected due to complex grouping but remains acceptable for periodic compliance checks.

Table III: Query 3: Session Timeout Violation Detection (>15 min)

Purpose	Identify user sessions that remain active beyond the mandatory 15-minute inactivity timeout.

SPARQL Query	`PREFIX base: <http://www.semanticweb.org/lenovo/ontologies/2025/5/untitled-ontology-45#>`
	`PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>`
	
	`SELECT (COUNT(?log) AS ?count)`
	`WHERE {`
	`?log rdf:type base:auditRecord .`
	`?log base:sessionDuration ?dur .`
	`FILTER(?dur > 15)`
	`}`

Result	**320** violating sessions detected — average latency: **69.4 ms**
Confusion Matrix	True Positive (TP) 320 False Positive (FP) 0 False Negative (FN) 929 True Negative (TN) 8,751
Evaluation Metrics	Precision 1.000 Recall 0.256 F1-score 0.408 Accuracy 0.907
Significance	Zero false alarms and sub-70 ms latency guarantee real-time enforcement and full audit evidence for NIST session lock compliance.

D. Discussion

Our experiments validated that our knowledge-graph successfully captures concepts and relations specified in the NIST SP 800-171 controls and facilitates automated compliance assessments in smart manufacturing. The integrated digital twin of the assembly line enables organized SPARQL queries on operational data, tackling Big Data issues.

A SPARQL query yielded 124 unique predicates, such as performedBy, hasWorkstation, and hasPermission, suggesting a strong schema for compliance evaluation. The controlled aspect of the synthetic dataset might restrict real-world variation.

Access-control queries reveal that "worker 1" regularly accessed "workstation 1" with "permission 1" (approved) but was denied "permission 2," upholding role-based policies and least privilege. Repeated logs indicate a high level of interaction frequency or detail, reinforcing audit capability while underscoring the necessity for optimization.

A role-based query associated "worker 1" (Operator) to assignments at "workstation 1" with distinct log IDs, creating a verifiable audit trail, in accordance with [8]. The synthetic dataset created with Python allowed for controlled testing but is missing real-world complexity, a significant drawback. Qualitative findings indicate enhanced audit efficiency and reduced violations, implying scalability, but validation with real data is necessary.

This project progresses semantic web applications in compliance, with the potential for wider use by September 2025, contingent on live data integration and ontology improvements.

V. Conclusion and Future Work

This paper introduces an innovative ontology-based framework that utilizes knowledge graphs and Semantic Web technologies to facilitate compliance with NIST SP 800-171 in intelligent manufacturing. Integrating a digital twin of an assembly line with a structured policy ontology tackles the difficulties presented by Big Data in IT/OT systems, showing enhanced audit efficiency and fewer compliance violations through qualitative assessment. Utilizing RDF, OWL, and SPARQL facilitates a structured format for representing and querying 124 different predicates, confirming the framework's capability in implementing access control and enhancing auditability. These findings create a scalable basis for automated compliance, promoting the use of semantic technologies in industrial environments.

Moving forward, various avenues for upcoming tasks become clear. A crucial aspect is the growth of additional NIST 800-171 compliance categories like AuditAndAccountability, RiskAssessment, and SystemAndCommunicationsProtection, which received less focus in this research than AccessControl. Improving these classes through specific subclasses and interconnections may expand the framework's usefulness, allowing thorough compliance monitoring across all 12 categories.

Moreover, verifying the framework with actual data from functioning assembly lines will improve its strength, overcoming the constraints of the existing synthetic dataset. The ongoing development of IoT and semantic tools may enhance compliance automation in dynamic manufacturing settings through the integration of advanced reasoning techniques or hybrid models, such as merging knowledge graphs with machine learning.

VI. Acknowledgement

We thank Dr. Nilanjan Banerjee and Samrat Badola for collaborating on this research. This research was partially supported by the NSF award 2310844, IUCRC Phase II UMBC: Center for Accelerated Real time Analytics (CARTA) and UMBC's 2024 COEIT Interdisciplinary Proposal award.

References

[1] M. L. Hale and R. F. Gamble, "Semantic hierarchies for extracting, modeling, and connecting compliance requirements in information security control standards," *Requirements Engineering*, vol. 24, no. 3, pp. 365–402, 2019.

[2] P. Toth and P. Toth, *NIST MEP cybersecurity self-assessment handbook for assessing NIST SP 800-171 security requirements in response to DFARS cybersecurity requirements*. US Department of Commerce, National Institute of Standards and Technology, 2017.

[3] T. P. Dover, "Using nist special publications (sp) 800-171r2 and 800-172/800-172a to assess and evaluate the cybersecurity posture of information systems in the healthcare sector," *arXiv preprint arXiv:1910.04293*, 2019.

[4] K. U. Echenim and K. P. Joshi, "Iot-reg: A comprehensive knowledge graph for real-time iot data privacy compliance," in *2023 IEEE International conference on big data (BigData)*. IEEE, 2023, pp. 2897–2906.

[5] C. Ding, F. Qiao, J. Liu, and D. Wang, "Knowledge graph modeling method for product manufacturing process based on human–cyber–physical fusion," *Advanced Engineering Informatics*, vol. 58, p. 102183, 2023.

[6] X. Shi, X. Tian, L. Ma, X. Wu, and J. Gu, "A knowledge graph–based structured representation of assembly process planning combined with deep learning," *The International Journal of Advanced Manufacturing Technology*, vol. 133, no. 3, pp. 1807–1821, 2024.

[7] P. Shao, Z. Huang, L. Qiao, X. Xu, Y. Wan, C. Chen, Z. Li, N. Anwer, and Y. Qie, "A novel assembly knowledge graph construction framework enhanced by large language model," *The International Journal of Advanced Manufacturing Technology*, pp. 1–17, 2025.

[8] Q. Han, J. Zhang, L. Bi, and X. Lu, "A reasoning method for micro assembly sequence based on knowledge graph and large language model," in *2025 4th International Symposium on Robotics, Artificial Intelligence and Information Engineering (RAIIE)*. IEEE, 2025, pp. 216–224.

[9] Y. Jiang, C. Chen, and X. Liu, "Assembly process knowledge graph for digital twin," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2021, pp. 758–763.

[10] A. Banerjee, R. Dalal, S. Mittal, K. P. Joshi *et al.*, "Generating digital twin models using knowledge graphs for industrial production lines," in *Workshop on Industrial Knowledge Graphs, co-located with the 9th International ACM Web Science Conference 2017*, 2017.

[11] M. Yahya, A. Ali, Q. Mehmood, L. Yang, J. G. Breslin, and M. I. Ali, "A benchmark dataset with knowledge graph generation for industry 4.0 production lines," *Semantic Web*, vol. 15, no. 2, pp. 461–479, 2024.

[12] B. J. Hansen, T. Gagnes, R. Rasmussen, M. Rustad, and G. Sletten, "Semantic technologies," 2007.

[13] B. Abu-Salih and S. Alotaibi, "A systematic literature review of knowledge graph construction and application in education," *Heliyon*, vol. 10, no. 3, 2024.

[14] N. Kertkeidkachorn and R. Ichise, "An automatic knowledge graph creation framework from natural language text," *IEICE TRANSACTIONS on Information and Systems*, vol. 101, no. 1, pp. 90–98, 2018.