Automated Knowledge Framework for IoT Cybersecurity Compliance

Ikechukwu Oranekwu
Dept. of Computer Information Systems
Texas A&M University - Central Texas
Killeen, USA
Ikechukwu.Oranekwu@tamuct.edu

Lavanya Elluri
Dept. of Computer Information Systems
Texas A&M University - Central Texas
Killeen, USA
elluri@tamuct.edu

Gunjan Batra
Dept. of Information Systems and Security
Kennesaw State University
Kennesaw, USA
gbatra@kennesaw.edu

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825755

***Abstract*—Rapid expansion in the manufacture and use of Internet of Things (IoT) devices has introduced significant challenges in ensuring compliance with cybersecurity standards. To protect user data and privacy, all organizations providing IoT devices must adhere to complex guidelines such as the National Institute of Standards and Technology Inter agency Report (NIST IR) 8259, which defines essential cybersecurity guidelines for IoT manufacturers. However, interpreting and applying these rules from these guidelines and the privacy policies remains a significant challenge for companies. Thus, this project presents a novel approach to extract knowledge from NIST 8259 for creating semantically rich ontology mappings. Our ontology captures key compliance rules, which are stored in a knowledge graph (KG) that allows organizations to crosscheck and update privacy policy documents with ease. The KG also enables real-time querying using SPARQL and offers a transparent view of regulatory adherence for IoT manufacturers and users. By automating the process of verifying cybersecurity compliance, the framework ensures that companies remain aligned with NIST standards, eliminating manual checks and reducing the risk of non-compliance. We also demonstrate that compared to the baseline Large Language Models (LLMs), our proposed framework has more compliance accuracy, and is more efficient and scalable.**

***Index Terms*—IoT, Cybersecurity, NIST 8259 standards, KGs, regulatory compliance, automated compliance, LLMs, privacy policies, SPARQL.**

I. Introduction

Internet of Things (IoT) is a network of connected devices that exchange data autonomously. As illustrated in Figure 1, IoT device connections have rapidly grown in the past decade. Moreover, according to [1], the anticipated growth in the number of active IoT devices is expected to surpass 25.4 billion by 2030. The proliferation of smart devices across various sectors and the large amount of data stored and processed by the IoT devices can be targeted by cybercriminals, who can gain unauthorized access to the data and use it for fraudulent activities. Cyberattacks on these devices can lead to breaches and loss of privacy. Therefore, it is critical to secure the IoT devices, especially due to the sensitive nature of this data. Regulatory authorities, particularly in the USA, have developed comprehensive cybersecurity standards and data protection regulations to safeguard users' information, such as federal IoT laws, the Cybersecurity Improvement Act [2], and Improving the Nation's Cybersecurity [3]. Also, it is essential to obey standards by the NIST primarily that apply to IoT device manufacturers like NISTIR 8259 [4] and NISTIR 8228 [5]. These regulations are encapsulated in complex documents, and as part of this research, we focus on NIST Standard 8259 [4], which outlines specific requirements for IoT device security.

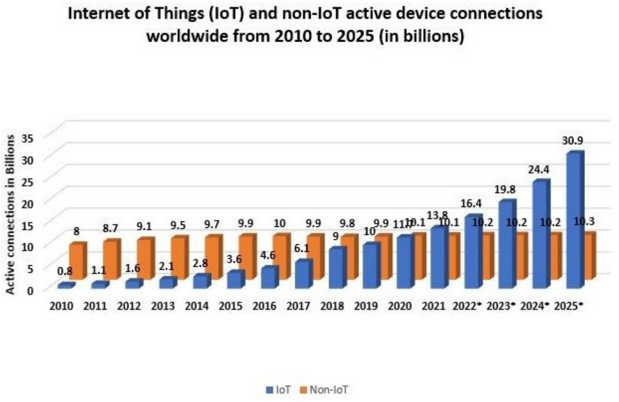


Fig. 1: Global IoT and non-IoT connections 2010-2025 [6]

For the IoT device manufacturing companies, ensuring compliance with these regulations is a daunting task, particularly when it comes to understanding and integrating the relevant rules into their privacy policies. Many companies struggle to interpret and implement these regulations effectively. The complexity of cybersecurity laws and the diversity of IoT devices make manual compliance checks difficult, error-prone and inefficient, thus highlighting the need for an automated system that can facilitate compliance checks. Hence, compliance with cybersecurity standards such as NIST 8259 is essential but a big challenge for IoT vendors.

Thus, with the expanding IoT ecosystem, automation is increasingly needed for cybersecurity compliance. Automated systems can quickly verify adherence to standards, reducing non-compliance risks. Leveraging technologies like KGs and LLMs can ensure timely and thorough compliance across IoT systems. A KG is a structured representation of entities and their relationships, enabling rich connections between data. It supports efficient querying and reasoning by modeling real-world concepts and their interactions. Additionally, KGs play

979-8-3503-6248-0/24/$31.00 ©2024 IEEE

a vital role in Artificial Intelligence (AI) by enabling semantic processing and open interconnection for intelligent services like search and personalized recommendations [7]. While there are attempts to use KGs in sectors such as finance and healthcare, their application in the domain of IoT cybersecurity compliance, particularly in automating the cross-verification of Business Requirement Documents (BRDs) against regulatory standards, has not been explored yet.

LLMs are AI systems trained on vast amounts of text to generate human-like language, answer questions, and assist with tasks like summarization or translation. Retrieval-augmented generation (RAG) combines LLMs with external data sources, retrieving relevant information during inference to enhance the model's accuracy and relevance.

In this work we present a novel approach to interpret and apply the guidelines in the policy documents. Essentially, we propose a framework that extracts knowledge from NIST 8259 and related IoT cybersecurity regulations, creating semantically rich ontology mappings. These ontologies capture key compliance rules, which are stored in a KG that allows organizations to crosscheck and update privacy policy documents with ease. The KG enables real-time querying of the system using SPARQL [8] and also offers a transparent view of regulatory adherence for IoT manufacturers and users. This framework automates the process of verifying cybersecurity compliance, thereby ensuring that companies remain aligned with NIST standards, eliminating manual checks and reducing the risk of non-compliance.

The key contributions of our work are as follows:

- Ontology-Based Knowledge Extraction: This structures NIST standards and data protection regulations into a KG.
- Integration with Semantic Web Technologies: SPARQL queries allow organizations to update privacy policies based on regulatory changes, ensuring continuous compliance with minimal manual effort.
- Automated Compliance Verification: Leverages LLMs to help automate compliance checks via the KG.
- Publishing KGs: It is publicly accessible and helps IoT manufacturers and users easily comprehend compliance without reading complex documents.

In this paper, in Section II we review related works on KG, LLMs, and IoT cybersecurity compliance. Next, in Section III, we present our methodology, detailing ontology development, LLM-based triple extraction, and KG construction. In Section IV we present a comparative analysis of the proposed KG-enhanced RAG system and experimentally evaluate the system's performance. Finally, in Section V, we conclude with suggestions for future work.

II. Related Work

A. Interoperability of Web of Things

The Web of Things (WoT) paradigm, introduced in the late 2000s, aims to address the interoperability challenge by leveraging web standards for the interconnection of embedded devices. This situation poses significant challenges regarding interoperability, and to mitigate these issues, semantic web technologies have been proposed as a viable solution to enrich raw IoT data, thereby facilitating better integration and communication among diverse IoT systems [1]. The fragmentation within the IoT landscape has intensified, necessitating systematic approaches to integrate web technologies into IoT scenarios [9]. Their literature presents a comprehensive taxonomy of WoT software architectures and enabling technologies. Furthermore, as the IoT ecosystem grows, the security of connected devices becomes increasingly critical. Another research [10], emphasizes the importance of defining the intended behavior of IoT devices to enhance cybersecurity measures. Manufacturer Usage Description (MUD) standard, established to describe the network behavioral profiles of IoT devices, offers a framework for understanding and managing potential security threats [10]. Despite its promise, the adoption of MUD in practical applications remains limited, indicating a need for further research to facilitate its implementation. Thus, the intersection of semantic technologies, web standards, and security frameworks presents a rich area of exploration for improving interoperability and enhancing the robustness of IoT systems. These studies underscore the necessity for ongoing collaboration between academia and industry to advance the state of IoT technologies and compliance.

B. NIST 8259

The National Institute of Standards and Technology (NIST) has addressed the security and privacy related challenges of IoT devices through publications like NISTIR 8259, which offers a framework for ensuring the security of IoT devices. NISTIR 8259 provides a baseline of recommended security features that manufacturers should implement during the design and production of IoT devices to minimize vulnerabilities [4]. In contrast to previous broader guidelines such as NIST SP 800-53 [11], which primarily focuses on security controls for federal systems, NISTIR 8259 is tailored specifically for the IoT environment. It emphasizes key areas like device identity, secure communications, and lifecycle management, addressing the unique challenges posed by the wide variety of IoT devices [4].

In this paper, we incorporate these guidelines into our ontology by representing NIST 8259 classes that map onto essential IoT security and privacy practices, particularly around manufacturer responsibilities. Next, we use KGs (which is a Semantic Web technology) [12]–[14] to aid in the management of this complex data. By embedding NISTIR 8259 within our ontology, we create a framework that not only addresses interoperability challenges but also enhances the overall security of IoT systems. This allows for more effective communication and integration of IoT device security standards into a wide range of deployment environments.

C. Knowledge Graphs in Regulatory Compliance

KGs provide a structured way to represent the relationships between various entities and rules within a domain, allowing for more effective querying and reasoning. Various studies

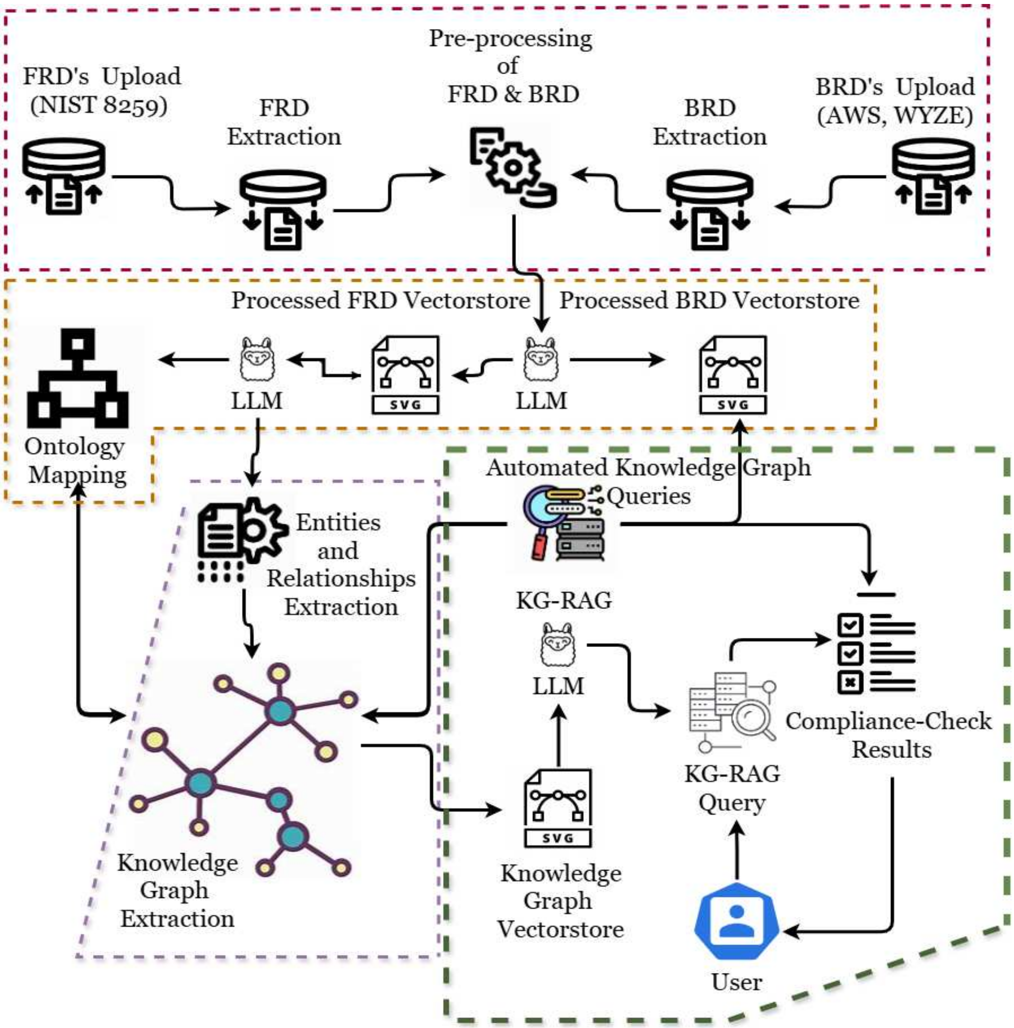

FRD's Upload (NIST 8259)
FRD Extraction
Pre-processing of FRD & BRD
BRD Extraction
BRD's Upload (AWS, WYZE)
Processed FRD Vectorstore
Processed BRD Vectorstore
LLM
SVG
LLM
SVG
SVG
Ontology Mapping
Automated Knowledge Graph Queries
Entities and Relationships Extraction
KG-RAG
LLM
Compliance-Check Results
KG-RAG Query
SVG
Knowledge Graph Extraction
Knowledge Graph Vectorstore
User

Fig. 2: Methodology Pipeline Overview

have shown the utility of KGs in sectors such as finance and healthcare, where regulatory compliance is critical. Some have also advocated for the leveraging of KG's databases for data compliance, however, the application of KGs in the domain of IoT cybersecurity compliance, particularly in automating the cross-verification of BRDs against regulatory standards, remains underexplored.

Several studies have highlighted the importance of automating compliance processes by leveraging KGs, which will allow for a structured representation of the intricate relationships between legal regulations and IoT security standards. For instance, Echenim et al. [15] developed IoT-Reg, a comprehensive ontology to ensure IoT data privacy compliance by integrating regulatory standards into a unified framework. The paper highlights how IoT-Reg assists manufacturers and users in understanding and adhering to regulations such as NISTIR 8228, HIPAA, and GDPR. Thus, IoT-Reg ontology aggregates privacy regulations and automates IoT data compliance checks through structured semantic relationships. Thus, KGs have emerged as powerful tools for managing and modeling complex regulatory environments. Previous research [16]–[20] introduced a semantically rich KG that integrates data compliance regulations in cloud environments, automating compliance for cloud service providers such as AWS and Google. Similarly, Kim et al. [21] developed a KG that automates HIPAA regulations for cloud-based health IT services, enabling healthcare organizations to maintain compliance with privacy rules through machine-processable formats. Further exploration of KG applications in unstructured document compliance was presented by [22], through their Deep Semantic Compliance Advisor (DSCA), which uses GNN-based models to compare complex contract clauses semantically. This approach has been particularly effective in domains like banking, where large volumes of unstructured data require interpretation. Chen et al. advanced the integration of KGs with LLMs to enhance decision-making in emergency management, demonstrating improved evidence-based decision-making through the structured knowledge of KGs [23].

The use of KGs for cyber defense exercises was examined by [24], who developed an ontology for managing data within cyber defense scenarios, enhancing integration and knowledge sharing in security contexts. Ontology Development for Cybersecurity Ontologies are essential for organizing domain knowledge in a structured and interoperable format, particularly in cybersecurity, where they can model threat detection, incident response, and compliance processes. Mozzaqua et al. [25] proposed an ontology-based cybersecurity framework for the IoT, which facilitates threat identification and dynamic security adaptation in real time. Ontologies are also critical in smart city applications, where IoT solutions are susceptible to cybersecurity risks. Andrade et al. [26] analyzed cybersecurity maturity in smart cities by proposing a model to assess risk levels and improve IoT cybersecurity practices. In the EU regulatory landscape, [27] reviewed the impact of evolving cybersecurity regulations on the IoT domain, particularly the EU's Cybersecurity Act and its implications for IoT supply chains. This work underlined the importance of ontologies and structured knowledge in automating compliance across sectors.

D. Retrieval-Augmented Generation (RAG) models with KGs

Retrieval-Augmented Generation (RAG) models combine LLMs with external knowledge sources like KGs to improve the generation of contextually relevant text and automate compliance checks. Xu et al. [28] proposed a novel dual-pathway approach that integrates KGs into RAG models to improve retrieval accuracy and mitigate hallucination during text generation. This method enhances the ability of RAG models to process domain-specific knowledge, making them more applicable to regulatory compliance tasks. In a similar vein, [29] demonstrated the integration of KGs with RAG models in customer service environments, showing how structured retrieval from a KG improves accuracy in answering customer queries. These studies underscore the potential of combining KGs with LLMs to address complex compliance challenges in highly regulated environments, including IoT cybersecurity. By integrating KGs into RAG models, organizations can leverage structured, domain-specific knowledge to automate and streamline compliance checks, making them more efficient and accurate.

III. Methodology

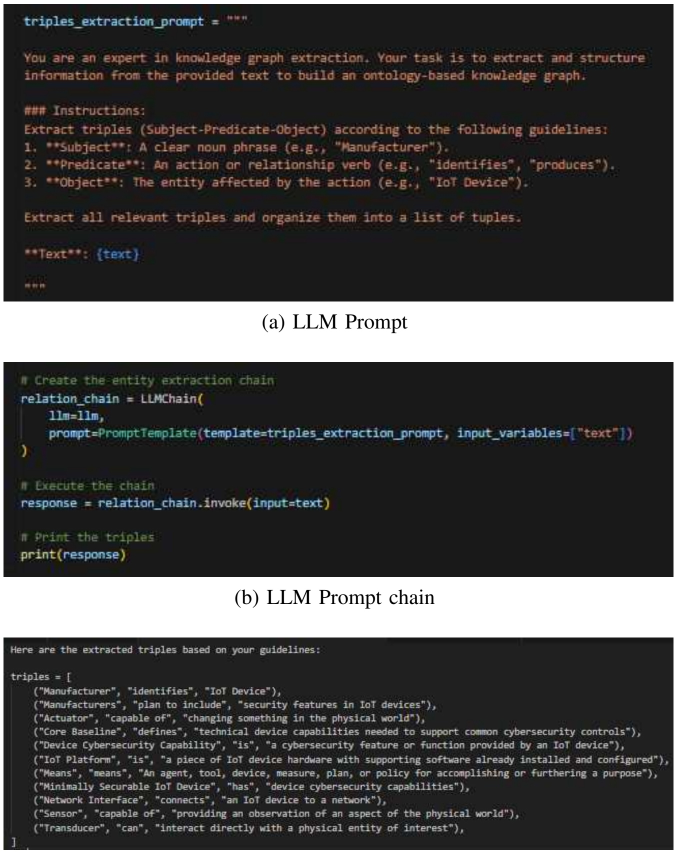
In this section, we present our methodology as illustrated in Figure 2, which involves four key stages:

1) Triple extraction,
2) Ontology development,
3) KG construction,
4) KG query and automated compliance verification using a KG-enhanced RAG model.

These stages collectively facilitate the extraction, structuring, and utilization of regulatory knowledge to verify IoT cybersecurity compliance.

A. Triple Extraction Using LLMs

LLMs have revolutionized natural language processing tasks, including text summarization, translation, and information extraction. Recent advancements have enabled LLMs to perform triple extraction, identifying subject-predicate-object relationships within text. These triples can then be directly mapped to an ontology, forming the basis of a KG. However, traditional LLM applications in compliance have been limited to document analysis rather than structured knowledge representation. Our work leverages LLMs to automate the extraction of triples from regulatory texts, a crucial step in building a compliance-focused KG. The extraction of subject-predicate-object triples from regulatory texts is performed using an LLM. We evaluate various LLMs to determine the most suitable for handling large volumes of data, balancing speed and accuracy. The LLM is guided by a prompt engineered to focus on identifying relationships relevant to IoT cybersecurity compliance. This automated process ensures that the KG is populated with accurate and relevant triples that reflect the content of the regulatory documents, as shown in Fig. 3. Effective Prompt Engineering for Triple Extraction—is crucial for maximizing the accuracy of triple extraction. The prompt is iteratively refined to improve the LLM's ability to identify entities and relationships that are critical for compliance verification. For example, prompts are designed to target specific sections of NIST standards that outline security requirements for IoT devices. The extracted triples are then validated to ensure their relevance and accuracy before being mapped to the ontology.



```
triples_extraction_prompt = """

You are an expert in knowledge graph extraction. Your task is to extract and structure
information from the provided text to build an ontology-based knowledge graph.

### Instructions:
Extract triples (Subject-Predicate-Object) according to the following guidelines:
1. **Subject**: A clear noun phrase (e.g., "Manufacturer").
2. **Predicate**: An action or relationship verb (e.g., "identifies", "produces").
3. **Object**: The entity affected by the action (e.g., "IoT Device").

Extract all relevant triples and organize them into a list of tuples.

**Text**: {text}

"""
```

(a) LLM Prompt

```
# Create the entity extraction chain
relation_chain = LLMChain(
    llm=llm,
    prompt=PromptTemplate(template=triples_extraction_prompt, input_variables=["text"])
)

# Execute the chain
response = relation_chain.invoke(input=text)

# Print the triples
print(response)
```

(b) LLM Prompt chain

```
Here are the extracted triples based on your guidelines:

triples = [
    ("Manufacturer", "identifies", "IoT Device"),
    ("Manufacturers", "plan to include", "security features in IoT devices"),
    ("Actuator", "capable of", "changing something in the physical world"),
    ("Core Baseline", "defines", "technical device capabilities needed to support common cybersecurity controls"),
    ("Device Cybersecurity Capability", "is", "a cybersecurity feature or function provided by an IoT device"),
    ("IoT Platform", "is", "a piece of IoT device hardware with supporting software already installed and configured"),
    ("Means", "means", "An agent, tool, device, measure, plan, or policy for accomplishing or furthering a purpose"),
    ("Minimally Securable IoT Device", "has", "device cybersecurity capabilities"),
    ("Network Interface", "connects", "an IoT device to a network"),
    ("Sensor", "capable of", "providing an observation of an aspect of the physical world"),
    ("Transducer", "can", "interact directly with a physical entity of interest"),
]
```

(c) Prompt Response

Fig. 3: Prompts

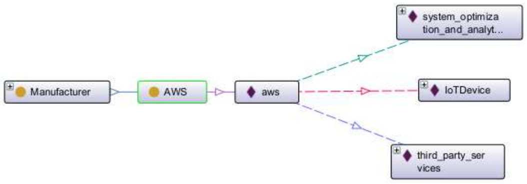


Fig. 4: AWS attributes (Manufacturer instance)

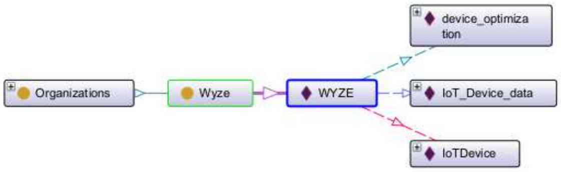


Fig. 5: Wyze attributes (Customer instance)

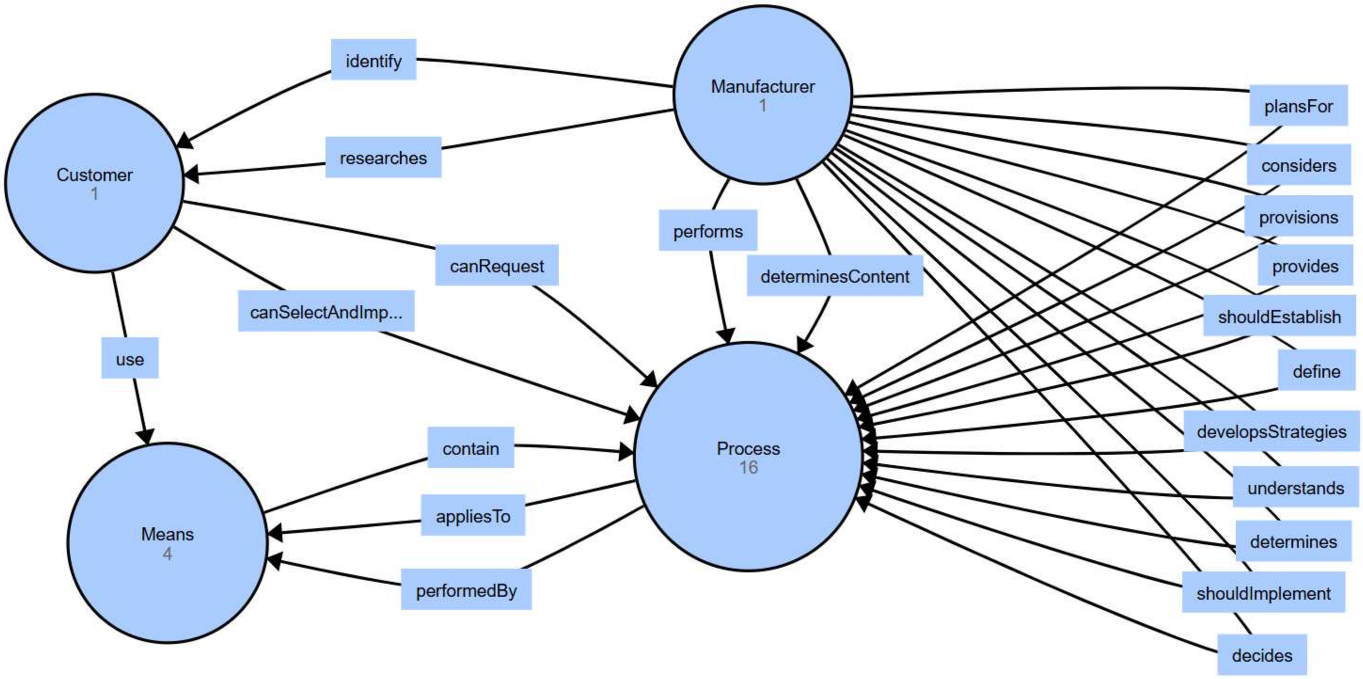


Fig. 6: NIST 8259 Entities and Relationships

B. *Ontology Development*

The ontology is the backbone of our KG, and our Domain-Specific Ontology Design (DSOD) and it captures the semantic relationships between entities defined in USA cybersecurity IoT laws, NIST standards (specifically NIST 8259), and other relevant regulations. It provides a structured representation of compliance processes, enabling more effective querying, reasoning, and automation for IoT cybersecurity compliance. Our ontology development process was heavily guided by LLMs using structured prompts to extract the relevant information from NIST 8259. The prompts were designed to capture:

1) Entities:- stakeholders, processes, and components central to IoT security. The key entities in NIST 8259 are:

- Stakeholders: represent the different actors involved in NIST 8259 such as IoT device manufacturers, users, and regulatory bodies like NIST. Each stakeholder has unique responsibilities and actions to ensure IoT cybersecurity compliance.
- Processes: represent actions or procedures related to managing, verifying, and ensuring IoT security. For example, configuration management ensures that IoT devices are configured securely, and compliance check represents the ongoing process of verifying adherence to regulatory standards.
- Means: represent the tools, mechanisms, or devices that are critical to the cybersecurity of IoT ecosystems. For instance, SecurityFeature refers to any component or feature that is implemented to enhance device security, while IoTDevice represents the actual devices that must comply with these standards.

2) Relationships:- connect these entities, reflecting the real-world responsibilities and interactions outlined in NIST 8259.
3) Data Properties:- defines the attributes of the entities, and key stakeholders relevant to NIST 8259, as shown in Fig. 4 and Fig. 5.
4) Object Properties:- define how entities interact with one another as represented in Fig. 6, ensuring that the relationships

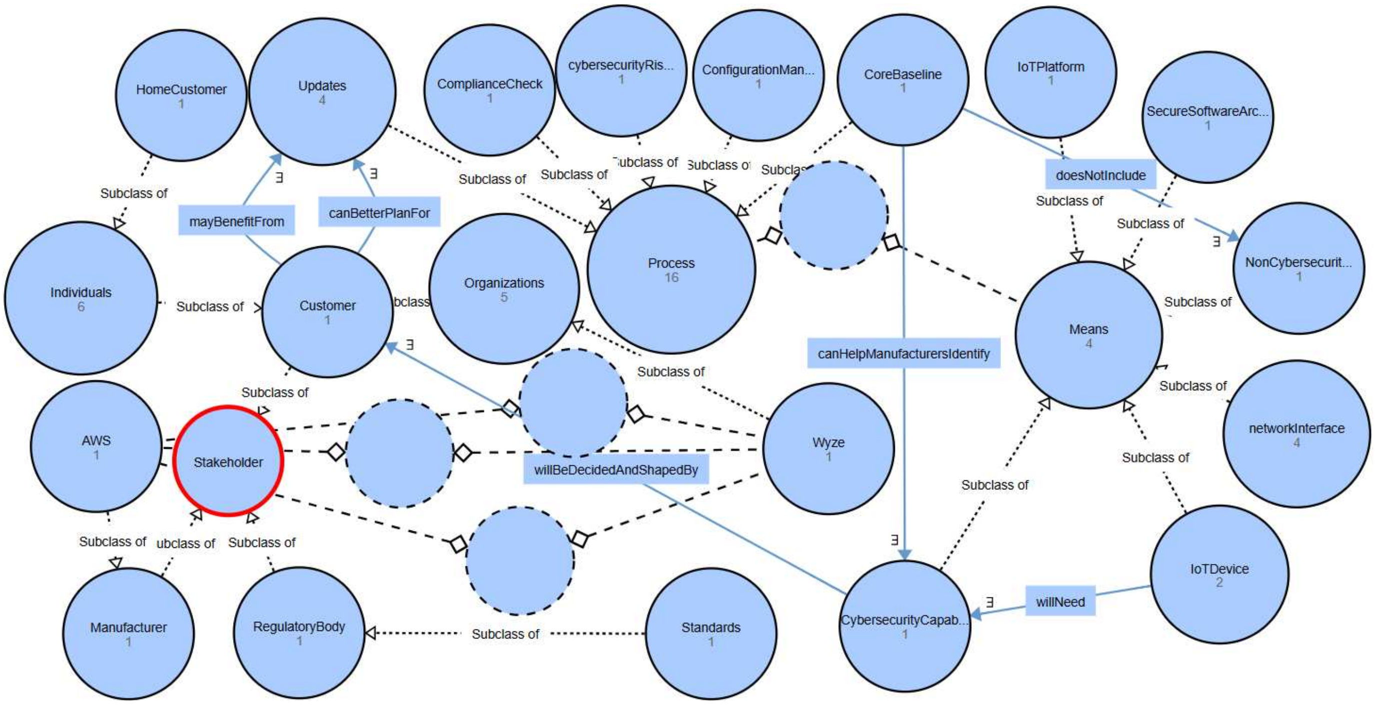
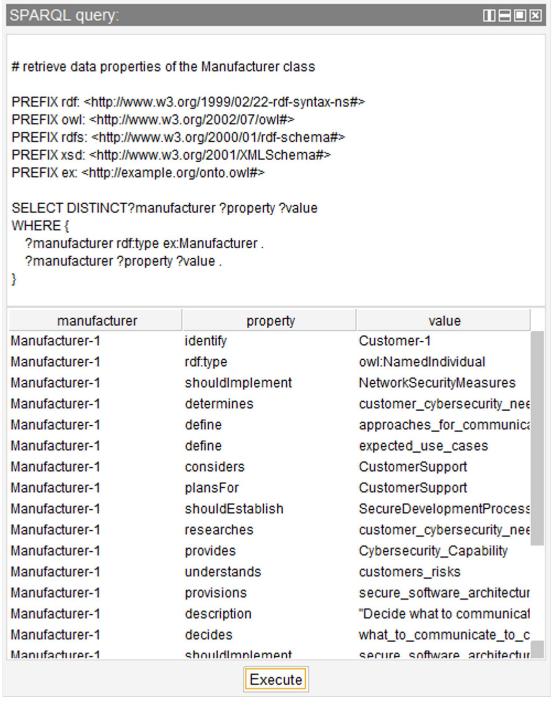


Fig. 7: NIST 8259 Knowledge Graph

outlined in NIST 8259 are accurately represented. Here are some key object properties:

- performs: defines the actions that a stakeholder, like a manufacturer, must carry out. For example, a Manufacturer may perform configuration management to ensure that IoT devices are securely set up. Thus, this property links Stakeholders like Manufacturer to Processes like configuration management, showing the responsibility of manufacturers to actively manage IoT security.
- isResponsibleFor: indicates responsibility for specific tasks or processes. For example, a RegulatoryBody such as NIST isResponsibleFor ensuring that standards like NIST 8259 are implemented and adhered to. Hence this property links Stakeholders like RegulatoryBody to Processes like compliance check, specifying regulatory oversight and enforcement.
- update: represents the process of updating IoT devices or their security measures over time. For example, a Manufacturer may update SecurityFeature as new threats emerge. This property connects Stakeholders (e.g., Manufacturer) to Means (e.g., SecurityFeature), indicating that manufacturers have a duty to continually enhance and update the security features of their devices.

Thus, our use of LLMs allowed for a semi-automated approach to building the ontology, speeding up the extraction of complex relationships from dense regulatory text. These extracted relationships were then formalized into an ontology framework that not only structures knowledge about the entities involved but also defines the interactions between them, to enable the KG serve as a reliable source for querying and reasoning about regulatory compliance.



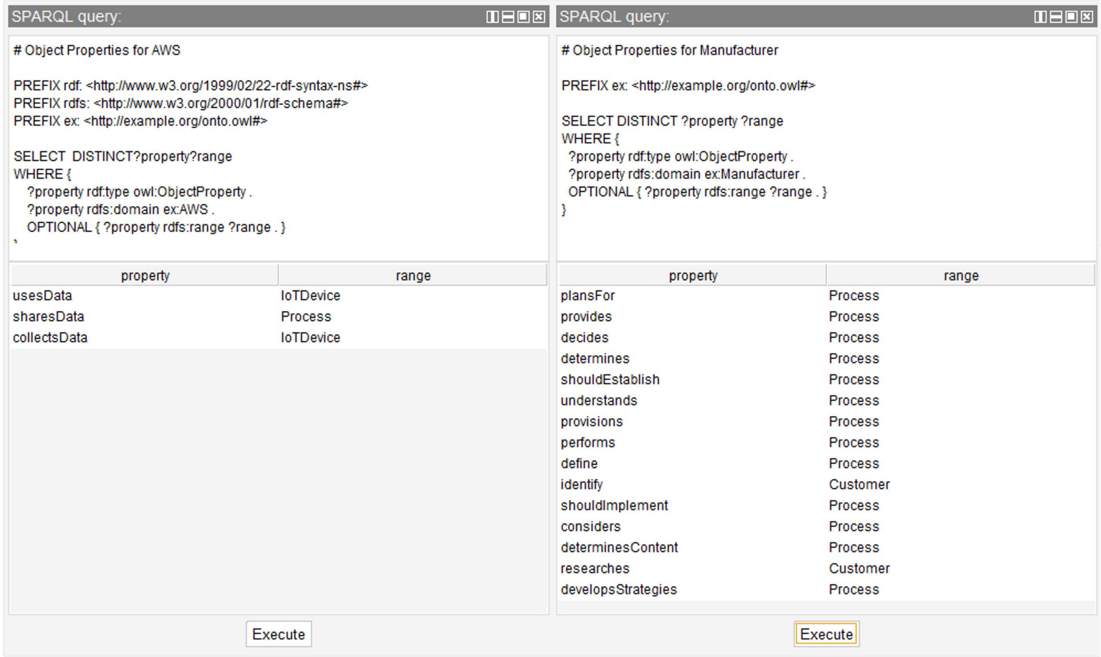
```
# retrieve data properties of the Manufacturer class

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ex: <http://example.org/onto.owl#>

SELECT DISTINCT?manufacturer ?property ?value
WHERE {
   ?manufacturer rdf:type ex:Manufacturer .
   ?manufacturer ?property ?value .
}
```

manufacturer	property	value
Manufacturer-1	identify	Customer-1
Manufacturer-1	rdf:type	owl:NamedIndividual
Manufacturer-1	shouldImplement	NetworkSecurityMeasures
Manufacturer-1	determines	customer_cybersecurity_nee
Manufacturer-1	define	approaches_for_communica
Manufacturer-1	define	expected_use_cases
Manufacturer-1	considers	CustomerSupport
Manufacturer-1	plansFor	CustomerSupport
Manufacturer-1	shouldEstablish	SecureDevelopmentProcess
Manufacturer-1	researches	customer_cybersecurity_nee
Manufacturer-1	provides	Cybersecurity_Capability
Manufacturer-1	understands	customers_risks
Manufacturer-1	provisions	secure_software_architectur
Manufacturer-1	description	"Decide what to communicat
Manufacturer-1	decides	what_to_communicate_to_c
Manufacturer-1	shouldImplement	secure_software_architectur

Fig. 8: Sample Manufacturer Query 1

C. *Knowledge Graph Construction*

Once the triples are extracted, they are mapped to the predefined ontology, forming a structured KG as depicted in Fig. (7) that captures the regulatory knowledge from the documents.



```
# Object Properties for AWS

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex: <http://example.org/onto.owl#>

SELECT  DISTINCT?property?range
WHERE {
   ?property rdf:type owl:ObjectProperty .
   ?property rdfs:domain ex:AWS .
   OPTIONAL { ?property rdfs:range ?range . }
`
```

property	range
usesData	IoTDevice
sharesData	Process
collectsData	IoTDevice

```
# Object Properties for Manufacturer

PREFIX ex: <http://example.org/onto.owl#>

SELECT DISTINCT ?property ?range
WHERE {
  ?property rdf:type owl:ObjectProperty .
  ?property rdfs:domain ex:Manufacturer .
  OPTIONAL { ?property rdfs:range ?range . }
}
```

property	range
plansFor	Process
provides	Process
decides	Process
determines	Process
shouldEstablish	Process
understands	Process
provisions	Process
performs	Process
define	Process
identify	Customer
shouldImplement	Process
considers	Process
determinesContent	Process
researches	Customer
developsStrategies	Process

Fig. 9: Sample Manufacturer Compliance check

This KG is then stored in a format compatible with Semantic Web technologies, OWL [30] or RDF [31] preferably, enabling efficient querying and reasoning. The construction of the KG involves aligning each triple with the appropriate classes and properties defined in the ontology, ensuring that the KG accurately reflects the regulatory landscape.

D. KG query and Compliance Verification using KG-Enhanced RAG

The final stages of the methodology involve querying the KG for compliance checks using SPARQL, then integrating the KG into an LLM with RAG capabilities. This integration allows the model to retrieve relevant regulatory information from the KG during the generation of compliance verification reports. By leveraging the structured knowledge in the KG, the RAG model can provide accurate and contextually relevant answers when cross-checking BRD against Functional Requirement Documents (FRD), as shown in Figures 10 and 8 respectively.

Eventually, we unit test an instance of manufacturer and their processes, querying the ontology to see if there is compliance, as shown in Fig. 9. Then we compare an instance of a customer and processes in their policy against the manufacturers to check for compliance as shown in Fig. 11.

RAG implementation and workflow help ensure seamless interaction with the KG. When a compliance check is requested, the model retrieves relevant triples from the KG, which are then used to inform the generative process. This workflow ensures that the outputs of the RAG model are not only accurate but also grounded in the structured regulatory knowledge stored in the KG. The result is a comprehensive compliance verification system that automates the process of cross-referencing BRD with applicable IoT cybersecurity standards.

IV. Experimental Evaluation

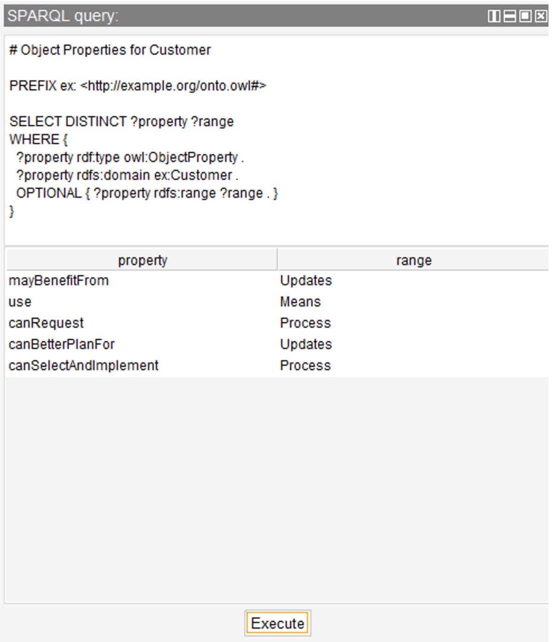
A. Experimental Setup

Our experimental setup involves testing the KG-enhanced RAG system on a dataset of BRDs and FRDs derived from real-world IoT cybersecurity scenarios. The documents used in the experiments cover a wide range of regulations, including NIST standards and USA data protection laws. The LLM is fine-tuned on domain-specific texts to ensure that it is well-equipped to handle the nuances of IoT cybersecurity compliance. The KG is constructed using the ontology developed in the earlier stages, and SPARQL queries are used to evaluate the system's performance in retrieving relevant regulatory information.

B. Performance Metrics

To assess the effectiveness of the KG-enhanced RAG system, we employ several performance metrics:

- Compliance Accuracy: The percentage of correct compliance checks performed by the system, indicating its ability to accurately cross-reference BRD with relevant regulations.
- Query Relevance: Measures the relevance of the information retrieved from the KG during the RAG process, ensuring that the outputs are contextually accurate.
- Efficiency: Evaluates the time taken to perform compliance checks, with the KG-enhanced system showing faster results due to the structured nature of the KG.
- Scalability: Assesses the system's ability to handle increasingly complex KGs without a significant drop in



```
# Object Properties for Customer

PREFIX ex: <http://example.org/onto.owl#>

SELECT DISTINCT ?property ?range
WHERE {
  ?property rdf:type owl:ObjectProperty .
  ?property rdfs:domain ex:Customer .
  OPTIONAL { ?property rdfs:range ?range . }
}
```

property	range
mayBenefitFrom	Updates
use	Means
canRequest	Process
canBetterPlanFor	Updates
canSelectAndImplement	Process

Fig. 10: Sample Customer Query 1

performance, ensuring that the methodology can adapt to future regulatory changes.

C. Results and Discussion

Here, we discuss our results, the quantitative and qualitative approaches we undertake to evaluate the impact of an enhanced KG + LLM pipeline for compliance verification. Our quantitative approach focuses on measuring the performance improvements in compliance accuracy and efficiency before and after KG integration. Then our qualitative approach examines the unique integration of querying and compliance verification features tailored to IoT device manufacturers, highlighting the advantages of using a KG built from regulatory documents.

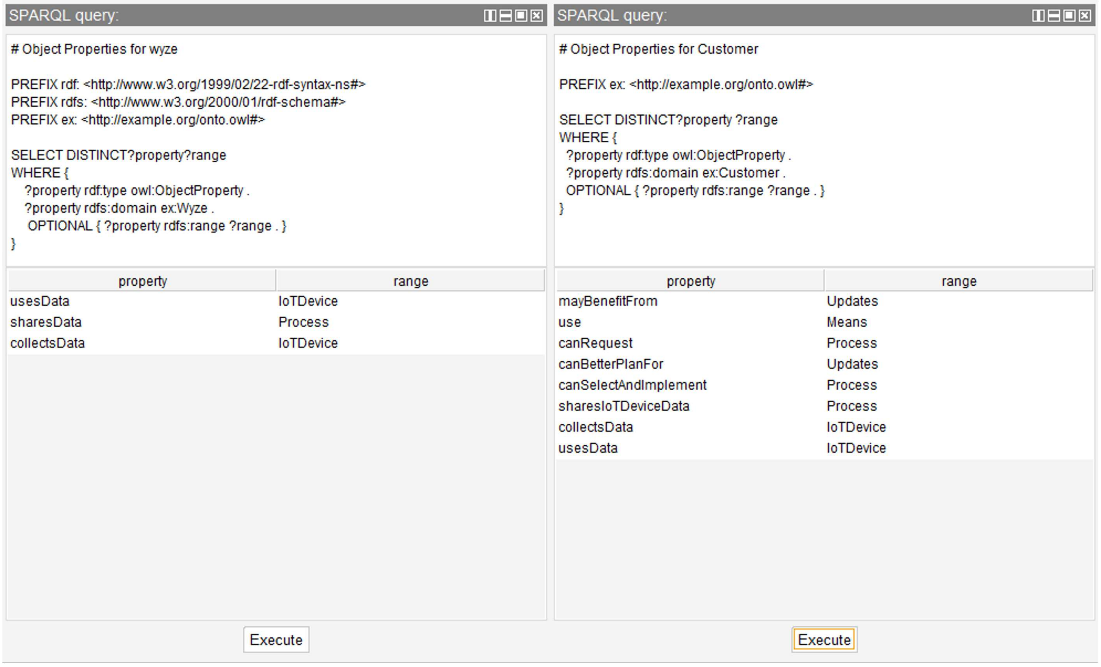
1) Quantitative Analysis Discussion:
To evaluate the impact of KG integration, we compare the performance of LLMs in compliance verification tasks Pre and Post KG + LLM integration. The baseline LLM, without KG support, relies solely on its pre-trained knowledge, leading to potential gaps in domain-specific understanding. Specifically, the KG-enhanced system achieved a compliance accuracy of 93%, compared to 75% for the baseline model. Query relevance also improved, with the KG-enhanced system consistently retrieving information that was directly applicable to the compliance checks. Thus, after integrating the KG, the LLM is able to retrieve precise regulatory information, significantly improving the accuracy and relevance of its compliance checks.
The efficiency of the system was evident in the reduced time required to perform compliance checks, highlighting the benefits of using structured data.
Scalability tests also showed the system could handle increasingly complex KGs without a significant drop in performance, making it a robust solution for ongoing compliance verification.
Overall, our analysis shows that the KG-enhanced RAG model outperforms the baseline LLM in terms of both compliance accuracy and efficiency. While KGs have been used in other domains such as finance and legal compliance, our approach is unique in its application to IoT cybersecurity. Existing systems often lack the flexibility and scalability needed to handle the rapidly evolving regulatory landscape of IoT. Our methodology addresses these challenges by leveraging LLMs for automated triple extraction and by designing an ontology that is specifically tailored to the intricacies of the adaptability and precision of IoT cybersecurity standards.

2) Qualitative Analysis Discussion:
To the best of our knowledge, this is the first work that not only creates a KG from the FRD (in our case NIST 8259 document) for IOT device companies, but also provides querying and compliance verification features to the users so that they can find gaps in their BRDs. There are no previous works that provide all these functionalities end-to-end to the IoT device manufacturers/companies. For instance, Echenim et. al [32] create IoT-Reg, (Internet of Things - Regulations), an ontology that encapsulates regulations and guidelines from NISTIR 8228, GDPR, and HIPAA and covers compliance and risk mitigation areas affecting various devices. While they have created the KG, we take the process further by adding the querying and compliance verification to our framework. Moreover, use of LLMs in our framework has enabled retrieval of precise regulatory information, significantly improving the accuracy and relevance of its compliance checks.

V. Conclusions and Future work

This paper presents a novel approach to automating IoT cybersecurity compliance checks by integrating a KG with an LLM using RAG. The methodology effectively addresses the challenges of understanding and applying complex regulatory standards by leveraging ontology-based knowledge extraction and structured querying. Our experimental results and case study demonstrate the significant improvements in compliance accuracy, query relevance, and efficiency provided by the KG-enhanced system. The system has demonstrated big data scalability and can handle complex KGs. A limitation of our work is that the accuracy of triple extraction is dependent on the quality of LLM and effectiveness of prompt engineering. Therefore, further optimization may be needed to ensure that performance remains consistent as the regulatory landscape evolves.

In the future, we will focus on scaling up the system to meet its big data potential, including optimizing data storage and retrieval mechanisms for even larger datasets. We plan to



```
# Object Properties for wyze

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex: <http://example.org/onto.owl#>

SELECT DISTINCT?property?range
WHERE {
   ?property rdf:type owl:ObjectProperty .
   ?property rdfs:domain ex:Wyze .
    OPTIONAL { ?property rdfs:range ?range . }
}
```

property	range
usesData	IoTDevice
sharesData	Process
collectsData	IoTDevice

```
# Object Properties for Customer

PREFIX ex: <http://example.org/onto.owl#>

SELECT DISTINCT?property ?range
WHERE {
  ?property rdf:type owl:ObjectProperty .
  ?property rdfs:domain ex:Customer .
  OPTIONAL { ?property rdfs:range ?range . }
}
```

property	range
mayBenefitFrom	Updates
use	Means
canRequest	Process
canBetterPlanFor	Updates
canSelectAndImplement	Process
sharesIoTDeviceData	Process
collectsData	IoTDevice
usesData	IoTDevice

Fig. 11: Sample Customer Compliance check

host and deploy the system publicly, providing open access to the KG and compliance verification tools.

Additionally, one potential area for future exploration is the development of effective prompting strategies for LLMs in the context of compliance verification. Given the complexity and specificity of regulatory language, optimizing how prompts are structured could lead to better performance in retrieving relevant information and interpreting compliance requirements. Techniques such as prompt engineering, few-shot prompting, and chain-of-thought prompting have shown promise in enhancing the reasoning and retrieval capabilities of LLMs, as highlighted in recent studies by Brown et al. [33], Wei et al. [34], Reynolds et al. [35] and Gao et al. [36]. Exploring these approaches in the context of IoT regulatory compliance could provide a pathway to further improve the accuracy and relevance of compliance checks. Thus, we plan to investigate ways to improve the LLM's reasoning capabilities over the KG, enabling more sophisticated compliance checks and broader applicability across different regulatory frameworks.

Acknowledgment

This work was supported by NSF award #2348147. We express our gratitude to colleagues whose insights and expertise significantly contributed to the research.

References

[1] F. Z. Amara, M. Hemam, M. Djezzar, and M. Maimor, "Semantic web and internet of things: Challenges, applications and perspectives," *ICOSI Laboratory, Khenchela, Algeria*, 2023.

[2] "Internet of things cybersecurity improvement act of 2020," Congress U.S. Government Publishing Office, Tech. Rep., 2020.

[3] T. W. House, "Improving the nation's cybersecurity," Executive Office of the President, Tech. Rep., 2021.

[4] J. T. Force, "Nistir 8259: Foundational cybersecurity activities for iot device manufacturers," 2020. [Online]. Available: https://doi.org/10.6028/NIST.IR.8259

[5] K. Boeckl, K. Boeckl, M. Fagan, W. Fisher, N. Lefkovitz, K. N. Megas, E. Nadeau, D. G. O'Rourke, B. Piccarreta, and K. Scarfone, *Considerations for managing Internet of Things (IoT) cybersecurity and privacy risks*. US Department of Commerce, National Institute of Standards and Technology . . . , 2019.

[6] L. S. Vailshery, "Bar chart for internet of things (iot) and non-iot active device connections worldwide from 2010 to 2025 (in billions) [chart]," Statista, 2022, global IoT and non-IoT connections 2010-2025. Available at: https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/.

[7] Z. Xu, Y. Sheng, L. He, and Y. Wang, "Review on knowledge graph techniques," *Dianzi Keji Daxue Xuebao/Journal of the University of Electronic Science and Technology of China*, 07 2016.

[8] W3C. (21 March, 2013) SPARQL 1.1 Overview. Accessed: August 23, 2023. [Online]. Available: https://www.w3.org/TR/sparql11-overview/

[9] L. Sciullo, L. Gigli, F. Montori, A. Trotta, and M. D. Felice, "A survey on the web of things," *Journal of Internet of Things*, vol. 1, pp. 1–10, 2022.

[10] J. L. Hernández-Ramos, S. N. Matheu, A. Feraudo, G. Baldini, and J. Bern, "Defining the behavior of iot devices through the mud standard: Review, challenges, and research directions," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3921–3935, 2021.

[11] J. T. Force, "Nist special publication 800-53 revision 5," 2020. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-53r5

[12] J. Hebeler, M. Fisher, R. Blace, and A. Perez-Lopez, *Semantic web programming*. John Wiley & Sons, 2011.

[13] G. Antoniou and F. Van Harmelen, *A semantic web primer*. MIT press, 2004.

[14] T. B. Passin, *Explorer's guide to the semantic web*. Manning Publications Co., 2004.

[15] K. Uzoma Echenim and K. P. Joshi, "Iot-reg: A comprehensive knowledge graph for real-time iot data privacy compliance," in *Proceedings of the 3rd Workshop on Knowledge Graphs and Big Data in Conjunction with IEEE BigData 2023*, 2023.

[16] K. Joshi, L. Elluri, and A. Nagar, "An integrated knowledge graph to automate cloud data compliance," *IEEE Access*, vol. 8, pp. 148 541–148 555, 2020.

[17] A. Kotal, L. Elluri, D. Gupta, V. Mandalapu, and A. Joshi, "Privacy-preserving data sharing in agriculture: Enforcing policy rules for secure

and confidential data synthesis," in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 5519–5528.

[18] R. Walid, K. P. Joshi, and L. Elluri, "Secure and privacy-compliant data sharing: An essential framework for healthcare organizations," in *International Conference on Mathematics and Computing*. Springer, 2024, pp. 15–26.

[19] L. Garza, L. Elluri, A. Kotal, A. Piplai, D. Gupta, and A. Joshi, "Privcomp-kg: Leveraging knowledge graph and large language models for privacy policy compliance verification," *arXiv preprint arXiv:2404.19744*, 2024.

[20] D.-y. Kim, L. Elluri, and K. P. Joshi, "Trusted compliance enforcement framework for sharing health big data," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 4715–4724.

[21] D.-Y. Kim and K. Joshi, "A semantically rich knowledge graph to automate hipaa regulations for cloud health it services," in *2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*. IEEE, 2021, pp. 7–12.

[22] H. Guo, B. An, Z. Guo, and Z. Su, "Deep semantic compliance advisor for unstructured document compliance checking," *IBM Research China*, 2021.

[23] M. Chen, Z. Tao, W. Tang, T. Qin, R. Yang, and C. Zhu, "Enhancing emergency decision-making with knowledge graphs and large language models," 2022.

[24] G. Babayeva, K. Maennel, and O. Maennel, "Building an ontology for cyber defence exercises," in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2022, pp. 423–432.

[25] B. A. Mozzaquatro, C. Agostinho, D. Goncalves, J. Martins, and R. Jardim-Goncalves, "An ontology-based cybersecurity framework for the internet of things," 2021.

[26] R. O. Andrade, L. Tello-Oquendo, S. G. Yoo, and I. Ortiz-Garcés, "A comprehensive study of the iot cybersecurity in smart cities," *IEEE Access*, 2021.

[27] P. G. Chiara, "The iot and the new eu cybersecurity regulatory landscape," 2021.

[28] S. Xu, M. Chen, and S. Chen, "Enhancing retrieval-augmented generation models with knowledge graphs: Innovative practices through a dual-pathway approach," 2021.

[29] Z. Xu, M. J. Cruz, M. Guevara, T. Wang, M. Deshpande, and X. Wang, "Retrieval-augmented generation with knowledge graphs for customer service question answering," 2022.

[30] W3C. (10 February, 2004) Web Ontology Language. Accessed: August 23, 2023. [Online]. Available: https://www.w3.org/TR/owl-features/

[31] W3C. (15 March, 2014) Resource Description Framework. Accessed: August 23, 2023. [Online]. Available: https://www.w3.org/RDF/

[32] 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.

[33] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[34] J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," *arXiv preprint arXiv:2201.11903*, 2022.

[35] L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," *arXiv preprint arXiv:2102.07350*, 2021.

[36] T. Gao, A. Fisch, and D. Chen, "Making pre-trained language models better few-shot learners," *Association for Computational Linguistics (ACL)*, pp. 3816–3830, 2021.