LLMQua:Using LLM automation to enhance risk quantification in data processing for big data platforms

1st Zhenyang Guo
the State Key Laboratory of I.S.N.
School of Cyber Engineering
Xidian University
Xi'an, Shaanxi
zyguo@stu.xidian.edu.cn

2nd Jin Cao
the State Key Laboratory of I.S.N.
School of Cyber Engineering
Xidian University
Xi'an, Shaanxi
jcao@xidian.edu.cn

3nd Wanying Ma
the State Key Laboratory of I.S.N.
School of Cyber Engineering
Xidian University
Xi'an, Shaanxi
wyma@stu.xidian.edu.cn

4rd Qiulan Xu
the Guangzhou Institute of Technology
Xidian University
Xi'an, Shaanxi
22151214540@stu.xidian.edu.cn

5th Ben Niu
Institute of Information Engineering
Chinese Academy of Sciences
Beijing, Beijing
niuben@iie.ac.cn

6th Hui Li
the State Key Laboratory of I.S.N.
School of Cyber Engineering
Xidian University
Xi'an, Shaanxi
lihui@mail.xidian.edu.cn

2025 11th IEEE International Conference on Privacy Computing and Data Security (PCDS) | 978-1-6654-7746-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/PCDS65695.2025.00054

***Abstract*—The compliant use of data has become an issue of significant concern for governments worldwide in the era of big data. Big data platforms span numerous fields, including education, government services, and e-commerce. Ensuring data compliance across these sectors has drawn considerable attention from both governmental and enterprise managers. To address this, relevant enterprises and scholars have developed numerous standards tailored specifically to various industry sectors. Thus, differentiating their similarities and differences demands extensive experience in platform development and in-depth knowledge of these standards. This implicitly elevates the difficulty and entry barrier for compliance detection in big data platform usage. The current standard compliance checks mainly rely on manual inspections, typically taking anywhere from one week to three months to complete. Recently, Zhan et al. proposed two automated methods enabling risk quantification for big data platforms based on standard documents. However, these approaches still have not resolved the automation of transforming standards into usable templates for these systems. In practice, developing such templates from standards requires experienced developers and typically takes from three days to one week, significantly restricting platform scalability. With the advancement of large language models (LLMs), their powerful text parsing capabilities have drawn extensive attention from researchers. Based on this, we propose a method named LLMQua, which utilizes LLMs to automatically generate risk quantification templates, reducing the standard-to-template conversion time to the scale of minutes. Experimental results show that the method achieves high accuracy and stability, and can be effectively integrated with the approach proposed by Zhan et al., significantly enhancing the scalability of the system.**

(*Corresponding author : Jin Cao.*), Email: jcao@xidian.edu.cn

This work is supported by the National Key R&D Program of China (No. 2022YFB3103400), and the National Natural Science Foundation of China (No. 62172317, U23B2024).

***Index Terms*—Large language models, Quantitative risk assessment, Big data platform.**

I. Introduction

With the ongoing wave of digitalization, the scale of big data assets and related investments continues to grow, driving sustained expansion in the big data industry. According to IDC's latest report, global big data IT investment reached approximately USD 293.2 billion in 2023 and is expected to approach USD 624.4 billion by 2028. Particularly in the Chinese market, IDC projects big data IT spending will reach USD 50.23 billion in 2028, accounting for around 8% of the global market, with the highest growth rate worldwide[1]. Driven by this significant growth, big data platforms have penetrated deeply into diverse sectors, such as power [1], medicine [2], finance [3], meteorology [4], and geology [5].

To ensure the secure operation and data compliance of these diverse big data platforms, countries such as China have proactively developed various security assessment models and protection requirements by bringing together enterprises and academic researchers from different fields. Additionally, the Chinese government has established dedicated departments, such as the Big Data Administration and the National Data Administration, and enacted relevant legislation, including the Data Security Law and the Personal Information Protection Law.

In comparison, the European Union began addressing these issues earlier, with the implementation of the General Data Protection Regulation (GDPR) dating back to 2018. Major

[1]https://www.idc.com/getdoc.jsp?containerId=prCHC52590324

978-1-6654-7746-8/25/$31.00 ©2025 IEEE
DOI 10.1109/PCDS65695.2025.00054

corporations such as Meta, Amazon, and TikTok have received substantial fines for violations. Consequently, to mitigate negative impacts arising from non-compliance, data enterprises typically conduct compliance inspections proactively. However, traditional compliance inspections rely heavily on manual labor and require experts with extensive industry-specific development experience and deep knowledge of relevant standards to verify compliance item by item, making this approach labor-intensive, time-consuming, and inadequate for large-scale and high-efficiency requirements. For example, a single self-assessment using the Data Security Capability Maturity Model typically takes about three months, and no automated solutions currently exist.

To address these challenges, rule-based, statistical, and machine learning methods have been gradually introduced to reduce the manual burden. However, these methods still require the construction of comprehensive risk models for accurate prediction. Ferreira et al. [6] developed a GDPR compliance checking system that enables external developers to embed compliance rules into the system, monitoring online platforms' data collection, usage, and protection. Although highly integrated, this solution mainly targets web applications developed using the MERN stack and requires significant adjustments to accommodate other frameworks or applications, thus limiting its applicability. Zhan et al. [7], [8] proposed RiskLen and its improved version, RiskTree. RiskLen utilizes random forest techniques to automate risk quantification assessments, reducing manual intervention with low computational costs. However, it relies heavily on expert scoring, uses limited data sources, and lacks validation in real-world scenarios. RiskTree enhances RiskLen by integrating knowledge graphs (KG) to automatically merge multi-source risk data, thereby improving accuracy, interpretability, and visualization capabilities, but at the expense of higher computational costs and stricter data quality requirements. Moreover, neither approach employs automated tools to extract key indicators directly from standard documentation. Consequently, if these methods [7], [8] are implemented solely based on China's national standard GB/T 35274-2017 Information Security Technology - Security Capability Requirements for Big Data Services, they will struggle to efficiently adapt to industry-specific big data platforms. System maintainers will face significant challenges in efficiently extending the system to other standard documents.

To enable more efficient adaptation of solutions such as [7], [8] to other standard documents, the key issue lies in converting these standards into templates usable by the solutions. Taking [8] as an example, the core concepts proposed are 'risk' and 'vulnerabilities', where the smallest functional unit of its system is defined as a 'vulnerability'. Detailed descriptions of the system proposed by [8] can be found in Section III. Hence, the aforementioned challenge can be framed as the task of converting key sentences from standardized texts into corresponding '*risks*' and '*vulnerabilities*', a typical Natural Language Processing (NLP) task.

In NLP, traditional methods, such as rule-based or template-based extraction techniques, possess high rule clarity and controllability, enabling precise extraction of structured information (e.g., clause numbers and key terms) while requiring relatively low computational resources [9]. However, these methods are limited by their heavy reliance on manually defined rules, making them ineffective at handling complex semantics or implicit logical relationships and leading to poor generalization. Frequent adjustments are necessary to accommodate textual variations across different standards [10]. In contrast, approaches based on Large Language Models (LLMs), such as GPT-4 [11], BERT [12], and DEEPSEEK [13], leverage deep semantic understanding gained from pre-training, enabling automated extraction of entities, relationships, and constraints from unstructured texts.

Based on the above considerations, this paper designs a system named LLMQua, which leverages LLMs to extract '*risks*' and '*vulnerabilities*' from relevant standards, and automatically adapts these extractions to systems such as [8], thereby providing efficient and automated standard-text adaptation capabilities. We summarize the primary contributions of this paper as follows:

1) We propose LLMQua, an automated risk and vulnerability extraction system based on fine-tuned LLMs, which acts as transformer, enabling systems such as [8] to efficiently adapt to new standard documents.
2) Experimental results demonstrate that LLMQua achieves a high accuracy of 99.83% and a stability score of 98.52%. Compared to manual methods, it significantly improves extraction efficiency, enabling the conversion of a single standard document within minutes.
3) Evaluations conducted on a telecommunications big data platform using the system from [8] indicate that standard templates generated by LLMQua effectively support automated risk quantification and identification.

The remainder of this paper is organized as follows: In Section II, we review related work on risk quantification and identification. In Section III, we revisit the RiskTree [8] approach. Section IV introduces the main workflow of LLMQua. Experiments and evaluations are presented in Section V. Finally, Section VI discusses the advantages of LLMQua and concludes the paper.

II. Related Work

Compliance detection is a critical challenge faced by enterprises globally. Under the influence of various regulatory policies worldwide, automating the conversion of standardized texts into quantification templates compatible with risk quantification systems is a valuable research area.

Traditionally, transforming compliance-related texts into machine-readable formats primarily relies on manual processing by experienced experts. This method is labor-intensive, time-consuming, and insufficient to meet large-scale, high-efficiency demands. In recent years, automated methods such as rule-based approaches, statistical methods, and machine learning have gradually been introduced. Bonatt et al. [14] developed a tool assisting data controllers and processors to

automatically verify if personal data handling and sharing comply with GDPR requirements. This tool identifies the structure of legal rules by thoroughly analyzing GDPR texts and employs the SPECIAL policy language to encode consent, business policies, and regulatory obligations into a machine-readable format for automated compliance verification. However, this method requires analyzing GDPR text structures and converting critical elements into machine-readable representations, presenting a barrier for non-specialists. Moreover, it lacks flexibility when applied to other regulations significantly different from GDPR in structure. Amaral et al. [14] assessed the compliance of privacy policies with GDPR's completeness criteria by calculating the similarity between individual sentences in the input privacy policy and sentences previously annotated with specific metadata types in a training set. However, this method neglects the contextual dependencies and logical structures inherent in legal texts. Aborujilah et al. [16] introduced NLP techniques in 2022, proposing an extended automated GDPR compliance verification tool that transforms legal requirements into natural language queries through a question-answering system, lowering the threshold for user comprehension. Nevertheless, this approach relies heavily on the MediaWiki API for document retrieval, meaning performance can degrade if resources become unavailable or outdated. Cejas et al. [17] presented an automated solution utilizing natural language processing techniques to evaluate the compliance of data processing agreements. This approach, however, depends on legal experts during the clause extraction phase, necessitating expert involvement whenever regulations change, thus increasing maintenance complexity and costs.

Compared to traditional methods, KGs enhance the semantic interconnectivity and interpretability of compliance rules by structuring legal provisions into organized data. For instance, Peng et al. [18] proposed an automatic code-compliance system leveraging Building Information Modeling and KGs, employing semantic web technologies to convert regulatory clauses into a KG. This formed a normative ontology that facilitates easier comprehension and identification of rules, thereby saving operational memory, execution time, and offering greater flexibility in data querying. However, this approach requires continuous effort and resources for constructing and maintaining KGs, and regulatory updates similarly increase maintenance costs. Echenim et al. [19] introduced a hierarchical KG based on IoT data lifecycles, integrating regulations such as NISTIR8228, HIPAA, and GDPR, yet the framework fails to account for data-flow changes resulting from device firmware upgrades, causing compliance analyses to lag behind actual business scenarios, notably in rapidly evolving fields such as cloud computing and IoT. Similarly, Chattoraj et al. [20] proposed the MedReg-KG framework, simplifying compliance detection by converting complex regulations into machine-readable KGs. Nevertheless, maintaining and updating these KGs still requires substantial ongoing effort and resources, increasing long-term maintenance costs. It is evident that such approaches also have limitations due to constantly evolving regulatory requirements and business environments. Therefore, dynamically adaptive knowledge bases are central to compliance detection. Traditional approaches require substantial labeled data, still facing difficulties in timely updates.

Given the challenges in dynamically updating and maintaining existing compliance KGs, innovative technical solutions from other domains provide valuable cross-domain insights for intelligent compliance risk management. For example, Chen et al. [21] proposed a novel framework for automated domain-specific KG construction, cleverly employing LLMs as domain experts. Using an iterative entity-induction tree-search algorithm, this approach achieves automation, specialization, and accuracy in KG construction, effectively reducing reliance on human intervention and thus broadening its applicability in practical scenarios. Similarly, Hu et al. [22] innovatively utilized LLMs to construct KGs from unstructured, open-source threat intelligence, leveraging GPT's few-shot learning capabilities for data annotation and augmentation. They subsequently created datasets for fine-tuning smaller language models, enabling topic classification, entity and relationship extraction, and the tactics, techniques, and procedures classification from collected reports. This method opens a new avenue for threat intelligence analysis and provides significant inspiration for analyzing unstructured legal and regulatory texts.

III. An Overview of RiskTree

In this section, we review the key definitions provided by Zhan et al. in RiskTree [8], and summarize the contributions of the RiskTree approach. RiskTree is an optimized risk quantification assessment method specifically designed for big data platforms, built upon the foundational RiskLens methodology. To address the shortcomings of traditional risk assessment methods in big data contexts, and overcome limitations in RiskLens such as insufficient granularity of indicators and reliance on subjective scoring for data acquisition, RiskTree reconstructs a more refined and systematic asset risk indicator framework and a data-processing risk indicator system. It innovatively integrates automated vulnerability scanning with dynamic questionnaire survey techniques, enabling comprehensive collection of multi-source heterogeneous risk data. Additionally, RiskTree employs KG (KG) technology for efficient integration and visualization of risk data. It also retains RiskLens's random forest algorithm, dynamically quantifying risk indicator weights and risk values through optimized risk calculation formulas. Thus, it achieves accurate risk assessment for both assets and data processing procedures. Experimental results demonstrate that, compared with RiskLens, the RiskTree approach is more objective, comprehensive, visually interpretable, and scalable, making it significantly more suitable for complex and dynamic big data security environments.

Methodology: RiskTree systematically optimizes existing risk quantification frameworks by categorizing big data platform assets into data, platform, and API assets, each with tailored risk indicators (e.g., unencrypted storage for data assets, weak passwords for platform assets, insufficient input filtering for API assets). It refines data-processing into nine phases

(access authentication, collection, transmission, provision, exchange, publishing, storage, backup/recovery, destruction) and precisely identifies phase-specific vulnerabilities, such as inadequate anonymization and unencrypted transmissions.

Moreover, RiskTree combines automated vulnerability scanning tools (e.g., Nmap, Hydra) with dynamically generated questionnaire surveys to capture both technical vulnerabilities and blind spots like network isolation checks. It employs KG to structurally connect assets, data-processing workflows, and vulnerabilities for intuitive risk visualization.

In quantification, RiskTree utilizes an optimized random forest model incorporating data sensitivity and asset-type weights. By dynamically linking vulnerabilities to threats via KG, it calculates comprehensive risk scores, delivering detailed, phase-specific risk assessments.

Advantages: RiskTree exhibits significant strengths in objectivity, interpretability, coverage, and scalability. Firstly, by integrating automated scanning tools and KG-based correlation analysis, RiskTree substantially reduces subjective reliance on expert scoring prevalent in traditional risk assessments, thus yielding more objective risk evaluation data. Secondly, KG-based visualization significantly enhances interpretability of risk assessment outcomes, allowing security professionals to intuitively understand inter-risk relationships and swiftly identify and respond to vulnerabilities. Additionally, RiskTree's detailed design of risk indicator systems for assets and data-processing workflows significantly expands the scope of risk scenarios, effectively identifying risks typically overlooked by traditional methods, such as compliance during data destruction and API access controls. Furthermore, the integration of questionnaire surveys and machine learning enables RiskTree to dynamically adapt to various application contexts, such as big data platforms in education and finance. Lastly, through automated vulnerability scanning and KG integration, RiskTree substantially reduces labor costs associated with risk assessments, enhancing the efficiency of risk quantification and effectively supporting large-scale platform assessments. Its modular design also provides flexible and convenient scalability for incorporating new risk indicators and adapting to new regulatory requirements in the future.

Disadvantages: Nevertheless, the RiskTree approach exhibits certain limitations. Firstly, it lacks sufficient automated processing capability for standard texts, currently requiring manual parsing and comprehension of standardized documents such as GB/T 35274-2017. This shortcoming makes automated extraction and mapping of unstructured standard texts difficult, thereby reducing the efficiency of compliance analysis. Secondly, the questionnaire survey system used by RiskTree has inherent constraints due to a limited pre-defined question bank, making dynamic adaptation to emerging standards challenging. This necessitates manual updating and expansion, preserving some degree of subjectivity. Furthermore, the binary response design (”yes” or ”no”) of the system fails to precisely capture more complex real-world security strategies, such as partial or localized data encryption, thus requiring further expert interpretation and confirmation. Finally, practical implementation of RiskTree requires experienced developers approximately 3 days to 1 week to construct usable standardized texts, significantly limiting the scalability of the platform. **Key Definitions:**

- **Risk:** A comprehensive measure indicating the probability and potential impact of negative incidents—such as data leakage, tampering, destruction, or service interruptions—that may arise due to vulnerabilities exploited by threats within the assets and data-processing procedures of big data platforms.
- **Vulnerabilities:** Implicit security defects or weaknesses objectively present within big data platforms as reflected in standardized texts. These vulnerabilities include, but are not limited to, deficiencies at the system, software, and management mechanism levels, which could be exploited by threats and consequently compromise data security, system stability, or service availability.

In summary, while RiskTree shows clear advantages in scalability, automation, and risk coverage, it faces significant limitations due to its insufficient automated processing of standardized texts. It heavily relies on manual interpretation by experienced personnel to convert unstructured provisions into structured templates, typically taking days to a week and restricting efficiency and scalability.

To overcome these limitations, we propose LLMQua, an innovative approach leveraging LLM fine-tuning techniques to automatically transform unstructured standard documents into structured risk templates directly usable by RiskTree. LLMQua substantially reduces manual intervention, enhancing processing efficiency and scalability, thus addressing RiskTree's critical deployment bottlenecks.

IV. The Proposed scheme: LLMQua

In this section, we present our proposed system, LLMQua, as illustrated in Fig.1. LLMQua focuses on extracting platform-relevant risks and vulnerabilities from unstructured standard documents in conjunction with platform-specific information, and subsequently generates ready-to-use questionnaires. Specifically, taking a typical RiskTree task as an example, the workflow can be divided into two phases: the preparation phase and the quantification phase. The preparation phase only needs to be executed once for a given standard document, whereas the quantification phase must be performed repeatedly.

As shown in Fig.2, to develop LLMQua, we first constructed a dataset derived from four prevailing Chinese-language standard documents. Two domain experts with extensive experience spent two weeks analyzing and decomposing these documents, resulting in a raw dataset comprising a total of 133 samples. To enhance system robustness, we applied data augmentation techniques such as synonym substitution and multilingual back translation, producing an expanded dataset referred to as *Dataset A*.

We then selected platform information from a telecommunications big data platform, transformed it into a specific JSON format, and automatically extracted a subset of platform

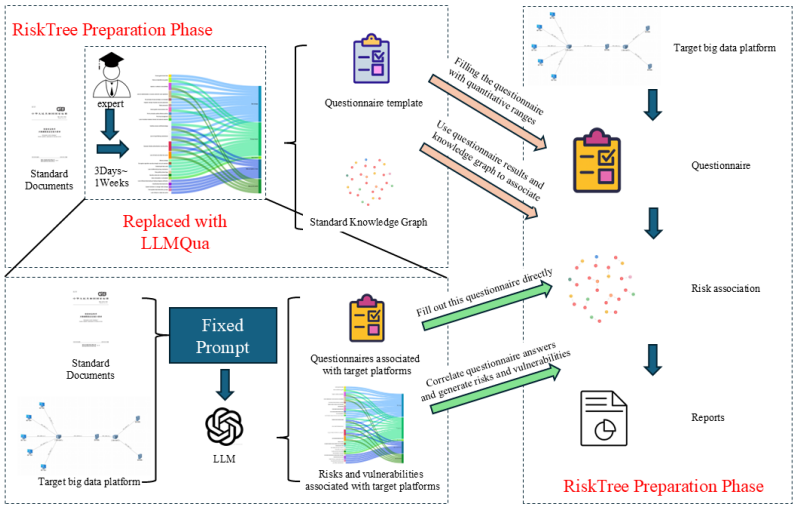

Fig. 1: Relationship between LLMQua system and RiskTree system

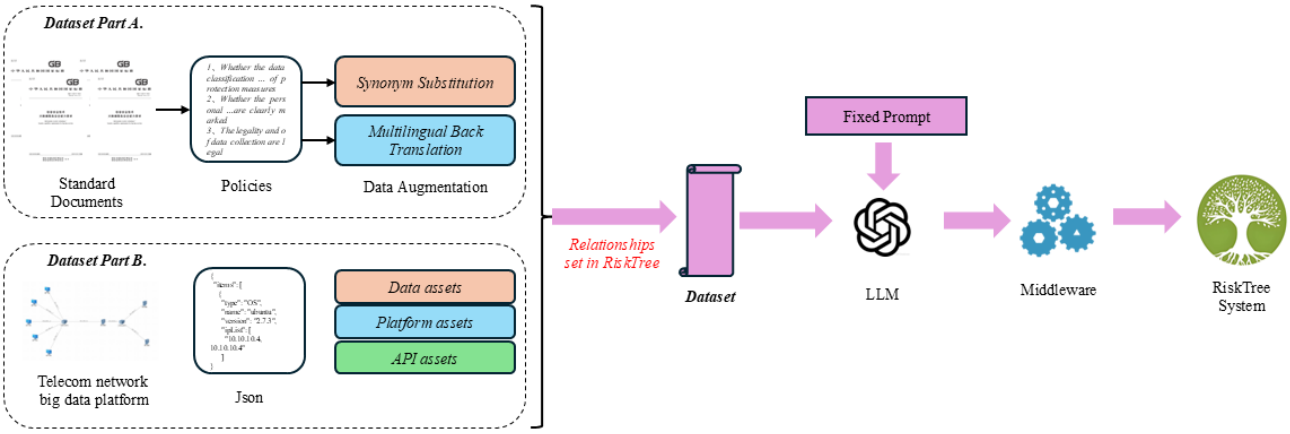

Fig. 2: The two-stage workflow of LLMQua, including the data preparation stage (left) and the model quantization and integration stage (right).

data. By randomly combining data assets, platform assets, and API assets, we constructed *Dataset B* in accordance with the asset–vulnerability relationships defined in RiskTree. This process resulted in a complete training dataset.

Subsequently, we fine-tuned several existing LLMs, including ChatGPT-4o, ChatGPT-4.5, DEEPSEEK, Kimi, and Doubao. An automated invocation tool was developed to complete the integration of RiskTree within the LLMQua framework.

A. Dataset construction

The dataset is primarily composed of two parts: Dataset A consists of risks and vulnerabilities defined in standards, along with their relationships, obtained through data augmentation techniques; Dataset B is constructed by freely combining subsets of real-world information extracted from a telecommunications big data platform. Ultimately, under the framework


```
{
  "Data breach":[
    {"Store data unencrypted":"Encrypted storage of sensitive personal information and important data and the effectiveness of encryption measures"}
    ...
  ]
  "...":[
    {"...":"..."}
    ...
  ]
}
```


Fig. 3: A sample of dataset A

defined by RiskTree, we integrated Dataset A and Dataset B to complete the construction of the full dataset.

Dataset A, as shown in Table I, was constructed by domain

TABLE I: Data Security Standards and Specifications

Standard Name	URL
Information security technology-Risk assessment method for data security	https://www.tc260.org.cn/file/2023-08-22/9702c85b-9c43-48f4-ac36-23021652f7be.pdf
Risk assessment specification for security of telecom network and Internet	https://hbba.sacinfo.org.cn/stdDetail/bfd8dc5e2b7ad0526255b1856ae6d27c8966a743957a1f0bcd8e7b8537181207
Big data platform security protection requirements for telecommunications and Internet	https://hbba.sacinfo.org.cn/stdDetail/ae32e5f1e1a5bb5022cdf00135cd5e6fa1753f821c0edec91db42e3fd77be511
Data security risk assessment implementation for telecommunications and Internet	https://hbba.sacinfo.org.cn/stdDetail/481da8f949af7f24fe05476dc8aab658128e5836f9bbca98f08b66d64583d9ec

experts who analyzed and decomposed four standardized documents. The resulting dataset follows the format illustrated in Fig. 3and includes a total of 133 unique, representative vulnerability samples as the original data. To enhance data diversity, we applied two data augmentation techniques—synonym substitution and multilingual back translation—ultimately expanding the dataset to 400 samples.

Dataset B describes the relationships among data assets, platform assets, and API assets in the platform using a JSON format, with IP addresses serving as the primary linkage. The telecommunications big data platform used in this study comprises nine servers and covers mainstream big data applications such as HDFS, Hive, and Flume. Through automated subset construction, we obtained 18 corresponding data samples.

B. LLM fine-tuning

Inspired by [23]–[25] et al., to leverage the advantages of current LLMs while reducing the overhead of deploying high-dimensional models locally, we adopt a task-specific instruction tuning approach using existing LLM APIs. Specifically, we design structured prompts to guide model behavior, and evaluate the models based on their performance under these instructions. Each instruction consists of four components: "Task Objective," "Expert Knowledge," "Platform Information," and the "Risk–Vulnerability Mapping" defined in RiskTree.

The *Task Objective* defines the goals we expect the LLM to accomplish. In the case of LLMQua, the model is required to perform two tasks: (1) extract "risks," "vulnerabilities," and their relationships from the input, using expert knowledge; and (2) generate questionnaires based on the extracted vulnerabilities, following the mapping rules defined in RiskTree.

The *Expert Knowledge* refers to a portion of the knowledge from Dataset A used in training, which explicitly defines how vulnerabilities, risks, and their relationships are represented in standardized texts.

The *Platform Information* corresponds to part of the knowledge from Dataset B used in training. It contains detailed platform context, such as which servers host which data assets, platform assets, and API assets.

Finally, the *Risk–Vulnerability Mapping* component reflects the relationships between vulnerabilities and platform assets as described by Zhan et al. [8] in the RiskTree paper.

Through this instruction-tuning setup, LLMQua fine-tunes LLMs to perform the two aforementioned tasks by learning from the three types of knowledge described above.

C. Middleware

The middleware component in LLMQua is primarily responsible for aligning the output of the LLM with the RiskTree system. This involves formatting and parsing the returned values from the LLM, writing the processed results into the database, and automatically updating the configuration files of the RiskTree system. The LLM is invoked using the OpenAI library, where the model output is accessed via MESSAGE.CONTENT. The extracted results are then converted into key-value pairs recognizable by the RiskTree system and used to automatically append new configuration entries in JSON or YAML format.

V. Experiment

A. Data Augmentation

Dataset A was constructed based on four industry standards through meticulous manual annotation, resulting in an original dataset containing 133 unique, representative vulnerability samples. Each sample is associated with 1 to 3 natural language descriptions, yielding a total of 169 descriptive texts. To enhance data richness, two data augmentation strategies—synonym substitution and multilingual back translation—were applied, ultimately expanding the dataset to 400 descriptive texts. The BERT model was used to compute the cosine similarity between augmented and original texts, enabling quantitative evaluation of the augmentation effectiveness. As shown in Fig. 4and Fig. 5 that the mean similarity for synonym substitution is 0.9765, for back translation is 0.8692, and for back translation after synonym substitution is 0.9391. Additionally, the similarity correlation between augmentation methods is low, with the correlation between synonym substitution and back translation as low as 0.04. These findings indicate that the augmented texts preserve semantic consistency while exhibiting methodological diversity. Systematic verification confirms that the augmented dataset effectively retains the semantic core of the original texts and introduces diversified representations, thus providing varied input for model training. The dataset has been validated to meet system requirements in terms of both quality and structural soundness.

B. Performance

In this section, we discuss the construction of LLMQua based on five state-of-the-art LLMs: Doubao-based, Kimi-based, DeepSeek-R1-based, ChatGPT-4o-based, and

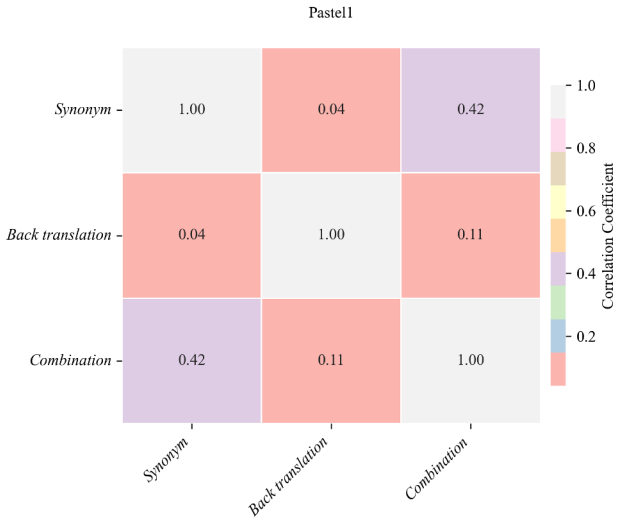

Fig. 4: Enhanced method similarity correlation heat map

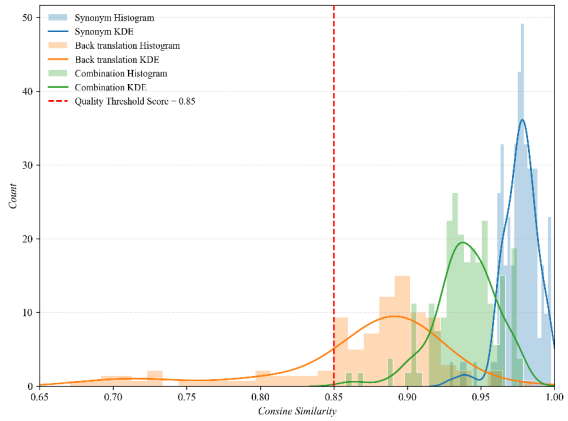

Fig. 5: Histogram of enhanced semantic similarity of text

ChatGPT-4.5-based versions of LLMQua. To evaluate these five configurations, we use two metrics: accuracy and stability.

1) Accuracy: For the accuracy metric, we account for the non-one-to-one relationship (NOTOR) between vulnerabilities and descriptive texts in Dataset A during the construction of the validation set. Based on this consideration, the performance of the five configurations varies, as shown in Table II.

TABLE II: Model Accuracy and Stability Comparison

Model	ACC. without NOTOR	ACC. with NOTOR	SI
Doubao-based	74.07%	94.28%	97.33%
ChatGPT-4o-based	58.15%	74%	98.07%
Kimi-based	81.48%	94.61%	96.15%
DeepSeek-R1-based	**96.30%**	**99.83%**	**98.52%**
ChatGPT-4.5-based	80.74%	80%	98.37%

The experiment result shows that DeepSeek-R1-based had the highest accuracy, reaching 99.83%. Doubao-based and Kimi-based also had high accuracy under the condition of removing duplicating items. ChatGPT-4.5-based showed a balanced performance under the two conditions, and the comparison of output results showed that it could better identify multiple vulnerabilities of the same text. However, there is a problem of insufficient accuracy, and the degree of differentiation of similar vulnerability names is not enough. The accuracy of ChatGPT-4o-based is slightly lower, which may be due to its poor performance in Chinese semantic analysis of correspondence. However, in summary, we can conclude that the method has a high accuracy in identifying vulnerability in a single corresponding text description, and also has a good performance in multiple vulnerabilities corresponding to the same text.

2) Stability: Stability is assessed from two perspectives: (1) comparing responses across different rounds of conversation for the same query to detect variations between earlier and later replies, and (2) examining consistency when repeating the same query within a single session or challenging previous answers by repeatedly asking for confirmation (e.g., "Are you sure this answer is correct?"). The degree of variation in responses is used as the basis for measuring stability, quantified by a stability indicator *SI*.

$$\mathrm{SI} = 1 - \frac{\sum_{i=1}^{N} \Delta A_i}{D \times R \times \sum_{j=1}^{M} T_j} \tag{1}$$

Here, ΔA_i denotes the number of content changes in the i-th query, D represents the total number of dialog instances, R is the number of test rounds per dialog, and $\sum T_j$ denotes the number of identified text items per round. A higher SI value indicates greater model stability and lower susceptibility to external prompting effects.

In our experimental setup, we set $R = 5$, conducting random repeated queries or challenge-based prompts in each round. Manual verification is performed to check whether the semantic meaning of the vulnerabilities has changed.

Experimental results show that the DeepSeek-R1-based, ChatGPT-4o-based, and ChatGPT-4.5-based versions of LLMQua achieve stability scores exceeding 98%. These models preserve original responses well under most challenge scenarios, indicating that the proposed method demonstrates high stability. The specific data are shown in Tab. II.

3) Energy consumption: Energy consumption is also an important part of measuring the merits of a method. We define energy consumption index EI to quantify energy consumption. The lower the energy consumption index, the lower the consumption and the higher the efficiency. We assume that all aforementioned LLMs operate under recommended configurations. For instance, the DeepSeek-R1 671B model requires at least 8 NVIDIA A100 GPUs (about 400W per GPU) when deployed using its recommended configuration. According to the energy consumption formula $E = P \times t$, the energy consumption per LLM and response latency per query are detailed in Table 3. In conclusion, the LLMQua

solution demonstrates lower energy consumption and significantly reduced processing time compared to manual processing approaches.

TABLE III: Model Energy Consumption Comparison

Model	Time (s/answer)	E (kj/answer)
Doubao-based	33.2	106.2
Kimi-based	35.2	112.64
DeepSeek-R1-based	28.8	92.16
ChatGPT-4o-based	57.2	183.04
ChatGPT-4.5-based	13.6	43.52

C. Application

By injecting the structured outputs generated by LLMQua into the RiskTree system, we executed the end-to-end risk quantification workflow. As illustrated in Fig. 6, the extracted knowledge was transformed into a knowledge graph, in which the central purple node indicates the adopted standard, anchoring all downstream risk associations. Orange nodes denote five major types of risks, blue nodes represent platform-relevant assets , red nodes are the vulnerabilities present in these assets, and beige nodes correspond to threats implied by those vulnerabilities. The graph structure enables automated propagation of risk relationships according to the logic defined in RiskTree. Due to space constraints, only a portion of the full graph is presented.

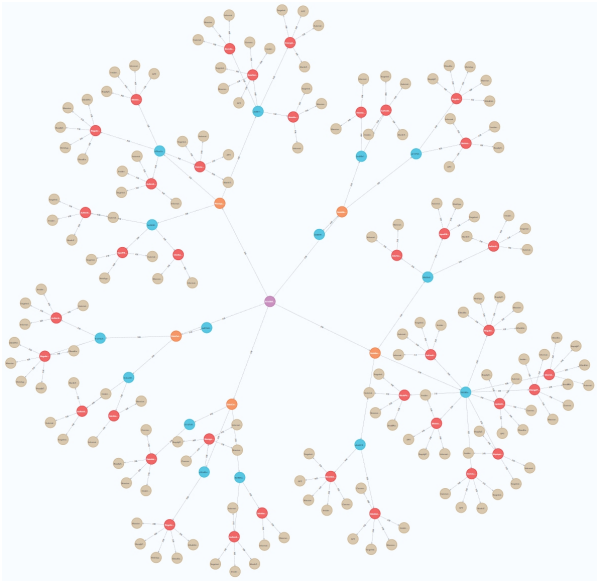

Fig. 6: Partial visualization of the knowledge graph generated by LLMQua and integrated into the RiskTree system. The central purple node denotes the name of the adopted standard, which anchors the risk evaluation process. Orange nodes represent the five predefined categories of risk types. Blue nodes are assets involved in the big data platform. Red nodes denote vulnerabilities identified in these assets. Beige nodes are the threats associated with each vulnerability. Directed edges indicate semantic relationships such as "has risk type," "possesses vulnerability," and "leads to threat," as defined in the RiskTree ontology. For readability, only a partial screenshot of the full graph is shown.

VI. Conclusion

In the era of big data, both governments and enterprises place significant emphasis on the compliant use of data. While the work by Zhan et al. has effectively addressed the automated identification of risks in big data platforms, it still fails to resolve the challenge of automatically converting standards into usable templates for these systems. To overcome the time-consuming, labor-intensive, and expertise-dependent limitations in developing system-specific templates from standards as observed in existing solutions, this paper proposes an LLM-powered approach for automatically generating risk quantification templates, thereby significantly enhancing the efficiency of standard conversion. Experimental results demonstrate that the proposed method can be effectively integrated with the framework developed by Zhan et al., substantially improving the scalability of the system.

References

[1] A.R. Al-Ali, R. Gupta, I. Zualkernan, and S.K. Das, "Role of IoT technologies in big data management systems: A review and Smart Grid case study," *Pervasive and Mobile Computing*, vol. 100, pp.101905 – 101934, 2024.

[2] X. Yang, K. X. Huang, D.W. Yang, W.L. Zhao, X.B. Zhou, "Biomedical big data technologies, applications, and challenges for precision medicine: a review," *Global Challenge*, vol. 8, pp.2300163 – 2300184, 2024.

[3] L. Theodorakopoulos, G. Thanasas, and C. Halkiopoulos, "Implications of big data in accounting: challenges and opportunities," *Emerging Science Journal*, vol.8, no. 3, pp. 1201–1214, 2024.

[4] J.Z. Chang, C. Gao, H.Y. Liu, F.N. Cui, D.S. Chang, and J.Cao, "Design and implementation of proxy server for meteorological big data platform Tianqing," in *Proceedings of Second International Conference on Big Data, Computational Intelligence, and Applications*, Huanggang, China, 2024, pp.1355002 – 1355009.

[5] M.T. Majemite, A. Obaigbena, M. A. Dada, J. S. Oliha, and P.W. Biu, "Evaluating the role of big data in us disaster mitigation and response: a geological and business perspective," *Engineering Science & Technology Journal*, vol.5, no. 2, pp. 338–357, 2024.

[6] M. Ferreira, T. Brito, J. F. Santos, and N. Santos, "RuleKeeper: GDPR-aware personal data compliance for web frameworks," in *Proceedings of 2023 IEEE Symposium on Security and Privacy*, San Francisco, USA, 2023, pp.2817–2834.

[7] H.M. Zhan,J.W. Yang, Z.Y. Guo, J. Cao, X.W. Zhao, W. You, and H. Li, "RiskLens: a novel way to quantify the risk for big data platform enhanced by machine learning," in *Proceedings of International Conference on Network Simulation and Evaluation*. Shenzhen, China, 2023, pp.228–242.

[8] H.M. Zhan,J.W. Yang, Z.Y. Guo, J. Cao, D. Zhang, X.W. Zhao, W. You, and H. Li, "RiskTree: Decision trees for asset and process risk assessment quantification in big data platforms," *Security and Safety*, vol.3, pp. 1–21, 2024.

[9] D.S. Asudani, N.K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," *Artificial intelligence review*, vol. 56, pp. 10345–10425, 2023.

[10] D. Nadeau, and S. Sekine, "A survey of named entity recognition and classification," *Lingvisticae Investigationes*, vol.30, pp. 3–26, 2007.

[11] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, l. Akkaya, and F.L. Aleman et.al, "GPT-4 Technical Report," 2023, arXiv:2303.08774. [Online].

Available: https://arxiv.org/abs/2303.08774
[12] J. Devlin, M.W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, Minnesota, USA, 2019, pp. 4171–4186.
[13] D.Y. Guo, D.J. Yang, H.W. Zhang, J. X. Song, R.Y. Zhang, and Q.H. Zhu et al., "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," 2025, arXiv:2501.12948v1. [Online]. Available: https://arxiv.org/abs/2501.12948
[14] P. A. Bonatti, S. Kirrane, I. M. Petrova, and L. Sauro, "Machine understandable policies and GDPR compliance checking,". *KI-Künstliche Intelligenz*, vol.34, pp.303–315, 2020.
[15] O. Amaral, S. Abualhaija, D. Torre, M. Sabetzadeh, and L. C. Briand, "AI-enabled automation for completeness checking of privacy policies," *IEEE Transactions on Software Engineering*, vol.48, pp. 4647–4674, 2021. O, S, D, et al. [J]. , 2021, 48(11): 4647-4674.
[16] A. Aborujilah, A. Z. Al-Othmani, Z. A. Long, N.S. Hussien, and D.A. Ghani, "Conceptual model for automating gdpr compliance verification using natural language approach," in *Proceedings of 2022 International Conference on Intelligent Technology, System and Service for Internet of Everything*, Hadhramaut, Yemen, 2022, pp.1–6.
[17] O. A. Cejas, M. I. Azeem, S. Abualhaija, and L. C. Briand, "Nlp-based automated compliance checking of data processing agreements against gdpr," *IEEE Transactions on Software Engineering*, vol.49, pp. 4282–4303, 2023.
[18] J. Peng, and X. Liu, "Automated code compliance checking research based on BIM and knowledge graph," *Scientific Reports*, vol.13, pp.7065–7077, 2023.
[19] K. U. Echenim, and K.P. Joshi, "Iot-reg: A comprehensive knowledge graph for real-time iot data privacy compliance," in *Proceedings of 2023 IEEE International Conference on Big Data*. Sorrento, Italy, 2023, pp. 2897–2906.
[20] S. Chattoraj, and K.P. Joshi, "MedReg-KG: KnowledgeGraph for Streamlining Medical Device Regulatory Compliance," in *Proceedings of 2024 IEEE International Conference on Big Data*, Washington, DC, USA, 2024, pp. 3382–3390.
[21] H. Chen, X. Shen, Q. Lv, J. Wang, X.Q. Ni, and J. P. Ye,"Sac-kg: Exploiting large language models as skilled automatic constructors for domain knowledge graphs," 2024, arXiv:2410.02811v1. [Online]. Available: https://arxiv.org/abs/2410.02811
[22] Y.L. Hu, F.T. Zou, J.J. Han, X. Sun, Y.L. and Wang,"LLM-TIKG: Threat intelligence knowledge graph construction utilizing large language model," *Computers & Security*, vol.14, pp. 103999–104010, 2024.
[23] J. Wei, M. Bosma, V.Y. Zhao, K. Guu, A. W. Yu, and B. Lester et al., "Finetuned language models are zero-shot learners,".2022, arXiv:2109.01652v5. [Online]. Available: https://arxiv.org/abs/2109.01652
[24] V.Sanh, A. Webson, C. Raffel, S.H. Bach, L. Suawika, Z. Alyafeai, and A. Chaffin et al., "Multitask Prompted Training Enables Zero-Shot Task Generalization," 2022, arXiv:2110.08207v3. [Online]. Available: https://arxiv.org/abs/2110.08207
[25] H.T. Xu, L.Y. Xin, Q.R. Yang, M. Li and M. Srivastava, "Penetrative AI: Making LLMs Comprehend the Physical World," in *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*, San Diego CA USA, 2024, pp. 1–7.