Compliance Compass (Automated Cybersecurity Adherence)

Azzam Turkistani Mohammed Alsayed Omar Dayili Osama Aldossary Salman Aldhalaan
KAUST Academy — CyS Division
Thuwal, Saudi Arabia

***Abstract*—Cybersecurity compliance presents significant challenges for organizations worldwide. As companies allocate substantial portions of their cybersecurity budgets to compliance efforts, manual processes have shown limitations and are susceptible to errors. The "Compliance Compass" project offers an automated solution that utilizes Large Language Models (LLMs) to generate comprehensive compliance reports. This system reduces the time and costs associated with compliance management while enhancing accuracy by prioritizing higher-impact standards in the compliance scoring process. Through an intuitive interface, organizations can effectively evaluate their compliance status, identify gaps, and receive actionable recommendations for improvement. The results demonstrate the system's ability to reduce manual effort and mitigate compliance risks, making it a valuable tool for strengthening an organization's cybersecurity posture. Future enhancements will focus on expanding the system's coverage of compliance standards and further refining its accuracy and adaptability.**

I. Introduction

In a world where data is under constant threat, the stakes in Saudi Arabia are especially high. The kingdom ranks 2nd globally in data breach costs, with an average of 38,800 compromised records per breach, far exceeding the global average of 25,575 [1]. The National Cybersecurity Authority (NCA) has published various standard documents and frameworks, each encompassing a distinct set of controls tailored to specific types of entities. Some controls are universally applicable, while others are designed for particular entities, such as educational institutions, online stores, and more. This has increased the resources needed to comply, hence, comes our project, the Compliance Compass.

II. Project Overview

A. Problem Statement

The increasing complexity of cybersecurity compliance presents significant challenges, with over 90% of organizations worldwide struggling to keep up with rapidly changing standards [2]. As Saudi Arabia advances its Vision 2030 goals, increased economic activity, digital transformation, and growing technology adoption have made the Kingdom a prime target for cyber attacks.

As Figure 1 shows, small companies that use automation were mostly compliant whereas those that did not were somewhat compliant or non-compliant. We can also deduce from Figure 1 that small companies that stopped using automation after EoQ1 had transferred to either somewhat compliant or non-compliant by the end of EoQ2.

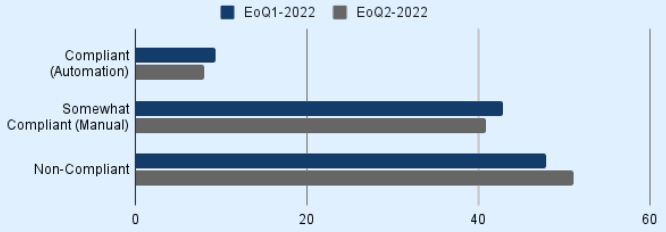


Fig. 1. Compliance status for organizations under 1,000 employees [3].

The reliance on manual processes further increases the risk of errors and non-compliance, leading to severe consequences such as legal penalties, data breaches, and reputational damage. With the cost of non-compliance often exceeding $14 million annually for many organizations [2], the financial impact highlights the need for a streamlined, automated solution.

B. Proposed Solution

The Compliance Compass automates the process of assessing the organization's compliance posture, reducing the time and cost needed for the assessment process. Organizations typically spend 40% of their cybersecurity budgets on compliance, but automation can lower these costs by up to 30% [4]. Using an LLM, the system will generate a compliance report based on a specific standards framework chosen by the company. The results will only be measured by relevant standards in the provided logs, with higher-impact standards weighted more heavily. The report identifies non-compliance areas and provides recommendations that allow organizations to maintain compliance with minimal effort. To generate the report, the LLM will be invoked with a detailed instruction created with a specific format to make the resulting report as accurate as it can be. The instruction format is mainly constructed from the type of the log, the 12 standards to compare against and their details, and a log file of a given application or device. This method allowed us to generate results without having to go through the costly fine-tuning process. Fine-tuning will probably generate more accurate and consistent results

but it is time-consuming and expensive especially in terms of computing power. Every company will have an account in the accounts database. This will allow each company to monitor their overall compliance posture over time.



Fig. 2. An abstract system architecture showing the main components.

III. Methodology

We recognized the need for a reliable automated system capable of measuring compliance with NCA regulations. To achieve this, our approach began with the preprocessing stage which included converting the data to a format that can be better understood by the LLM. This was critical in analyzing the data to identify potential compliance gaps and ensure alignment with the chosen regulatory requirements. Additionally, to ensure robust functionality, we developed the back-end using the OpenAI API.

As for the model selection, the choice to utilize gpt4o-mini was influenced by our need for an optimal balance between performance and accuracy. By "performance and accuracy," we refer to the model's capability to perform operations swiftly while delivering precise analysis. The gpt4o-mini is more compact compared to its larger counterpart, gpt4o, allowing for faster and far cheaper execution. Its reduced resource consumption aligns perfectly with our program's requirement for continuous operation and efficient processing of large log files.

This method enabled us to create a powerful processing engine that handles data, processes compliance checks, and manages report generation. The back-end is designed to process large volumes of data efficiently, ensuring secure data handling, and providing real-time integration with the LLM for accurate compliance assessments. This architecture ensures that the system can handle the complex data found in the logs while maintaining high performance and reliability.

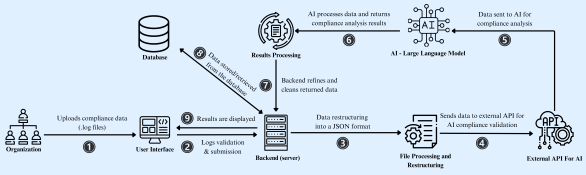


Fig. 3. The system workflow showing the steps of report generation.

The process involves several steps. Firstly, organizations begin by securely uploading their log files through the web interface. These files are then converted to a structured format as we stated above. Secondly, the LLM processes the data, comparing it against established compliance regulations, and the system generates a detailed compliance report based on this analysis.

We opted to use the JSON structure for converting the organization's log files and processing and producing the outputs from the LLM. JSON's flexibility facilitates seamless data transfer and interaction across different systems, presenting data in an organized manner that is easy for diverse systems to interpret. As the structure favored by LLM for data transmission, JSON's compatibility with most programming languages make it an exemplary choice for integrating disparate systems.

To store the results processed by the LLM model, we chose a database solution, which ensures data organization and swift access when needed. This method not only guarantees data security but also restricts access exclusively to the authorized organization, thereby safeguarding sensitive information.

Developing the LLM's System Instructions began with a set of generalized guidelines that were progressively refined to become more precise and specific. This iterative process involved analyzing the results and providing feedback, describing the scope of data processing, and incorporating exceptions when necessary. Such modifications aim to enhance accuracy and mitigate potential errors.

IV. Results

To delineate the outcomes derived from the rigorous assessments conducted via the Compliance Compass, we have produced the following results that provide insights into the system's efficacy in standard compliance evaluation, accuracy, consistency, token utilization, and error management.

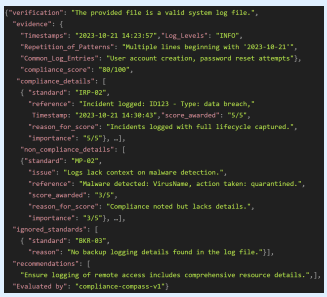


Fig. 4. A Snippet of the Raw Output from the API.

The sample output, seen in Figure 4, provides detailed metadata, including the status of the file verification test, and its associated evidence. It also elaborates on the score, compliant standards, non-compliant standards, and ignored standards. Along with those are the system-generated recommendations.

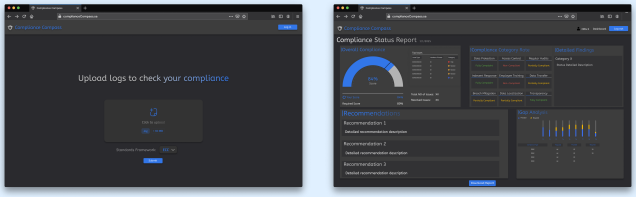


Fig. 5. Evaluation Report Details Represented on the User Interface.

The initial screen features an intuitive drag-and-drop interface complemented by a drop-down menu, allowing users to select their desired set of standards for compliance assessment. The evaluation report itself comprises four distinct sections, as shown on the right image in Figure 5. The first section provides an aggregate score that summarizes the extent to which compliance standards are met, offering an overall view of organizational adherence. Diving deeper, the second section offers a more thorough compliance status for each of the standards from the selected set (in this case, the 12 selected standards from the ECC). The third section

presents system-generated recommendations prioritized by importance from the most critical to the least essential, guiding users on the most impactful adjustments. Finally, the gap analysis diagram in the fourth section visually represents the work needed to achieve full compliance, highlighting existing disparities between the current state and the desired compliance status.

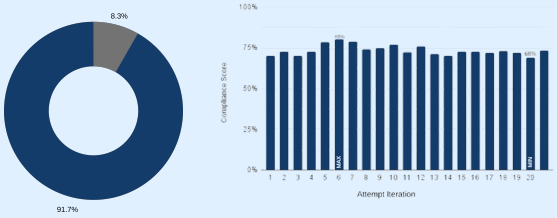


Fig. 6. Accuracy and Compliance Scores Per Log File.

In our accuracy assessment (seen in Figure 6), we executed 46 experiments, each against the 12 selected standards. Of the resulting 552 evaluated standards, a mere 12 were ignored, depicted in the donut chart in Figure 6. Among these, only one instance was inaccurately ignored, while the remaining 11 were correctly ignored.

For consistency evaluation, we carried out 20 tests on an identical log file. The scores ranged, with the highest being 80% and the lowest at 68%, culminating in an average score of 73.42

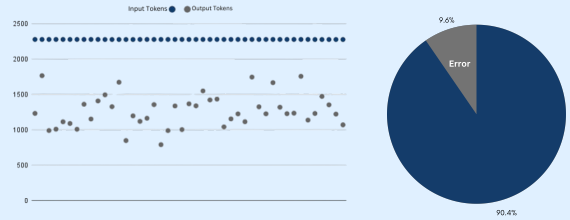


Fig. 7. Token Utilization and Error Rate.

In evaluating the error rate, we ran 66 experiments across 9 log files, each tested five times, other than for one instance. This is shown in Figure 7. Errors were identified as any output that was incomprehensible or nonsensical.

The assessment of token utilization was conducted similarly to the accuracy test. We performed 46 trials on 9 different log files. The input tokens remained stable at approximately 2500 tokens across all runs. Outputs varied from 500 to 2000 tokens, with total tokens expanded per test not exceeding 4500.

V. Analysis

In Figure 4, the system output has proven highly effective for front-end processing due to the JSON's organized structure. The output contains several primary sections such as the compliance overall score, compliance details, non-compliance details, the ignored standards, and recommendations to enhance the company's overall compliance. In Figure 5, the 1st UI screen will be used to upload the company's log files and the user will choose the standard framework to assess against from a drop-down menu. The 2nd UI screen contains 4 main sections. The 1st section shows an overall compliance score (Of 100) that is based on the total of each standard score. Additionally, the section contains the top issues part and it is color-coded to emphasize the significance grade of each issue. In the second section, we have the rate of each compliance category(Fully, Partially, or non-compliant) these rates will be given based on the score of the category. The detailed finding part will show category details such as which events in the log were relevant to each category. The 3rd section is the recommendations, it will show some points that help the company increase its compliance score and improve its overall compliance posture. Finally, the gap analysis section will provide an idea of what each compliance category is lacking. (Figure 6) Reflects the strong accuracy in correctly identifying only relevant standards from each log. The minor anomaly stemmed from the system's instruction flow, which included repetitive "abort" commands, thus possibly confusing the model. Additionally, the system's score variance is promising. Not to mention, this variance is primarily attributed to service-related issues, such as inconsistent token counts in identical runs, and would be reduced if these inconsistencies were addressed. In (Figure 7) the errors were mainly due to prompt engineering flaws. While this error rate is acceptable for initial deployment, it suggests areas for further refinement. The token usage chart demonstrates excellent scalability and efficiency. It also shows an excellent scalability and cost-effectiveness, operating within a 128k context window, with current logs averaging 5k tokens. This is particularly advantageous following the introduction of the latest large language models (LLMs).

VI. Future Work

Future improvements to the Compliance Compass will include expanding its coverage to include additional critical frameworks, such as SAMA. To achieve higher levels of accuracy and efficiency, we plan to fine-tune an open-source LLM using KAUST's Shaheen III supercomputer. Fine-tuning is essential because it will allow the model to be customized specifically to the requirements of different frameworks, improving its ability to identify relevant standards in log files and reduce false positives. Leveraging the supercomputer will enable the system to handle larger datasets and more complex scenarios with greater precision, reducing randomness and enhancing the relevance of the results.

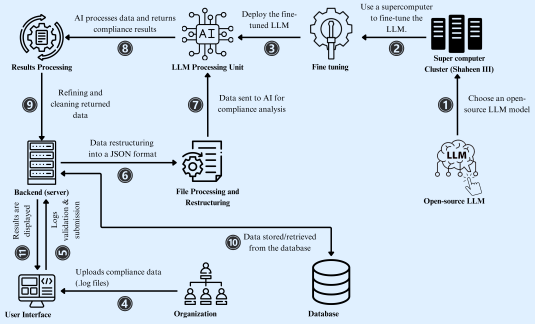


Fig. 8. A visualization of the future system's workflow after fine-tuning the LLM.

In addition to fine-tuning, another possible modification involves developing a smaller and more specialized model. This model would be specifically designed for compliance assessments, with a narrower focus and optimized performance for logs and standards analysis. Such a model could be more efficient in terms of both processing power and speed, allowing for quicker compliance checks while maintaining high accuracy. This would

be especially useful for organizations with limited resources, as it would reduce the operational expenses associated with large-scale, general-purpose models.

VII. Conclusion

To summarize, Compliance Compass exhibits strong performance in analyzing log files for compliance assessment. It is accurate, scalable, and cost-effective, making it a feasible solution for real-world applications. The Compliance Compass also shows a high potential for expandability and wide coverage of business domains by incorporating additional standard frameworks. This approach contributes to having a solution that reinforces cybersecurity in the kingdom and therefore aims to make a safe and reliable cyberspace which is one of the NCA 2030 vision goals.

References

[1] "Cybersecurity and its impact on digital saudi," Independent Directors Council, Tech. Rep., 2019. [Online]. Available: https://resources.trendmicro.com/rs/945-CXD-062/images/Cybersecurity-and-its-Impact-on-Digital-Saudi.pdf

[2] P. Institute, "The true cost of compliance with data protection," Globalscape, Tech. Rep., 2017. [Online]. Available: https://static.fortra.com/globalscape/pdfs/guides/gs-true-cost-of-compliance-data-protection-regulations-gd.pdf

[3] "State of ccpa and gdpr privacy rights compliance research report – q2 2022," Cytrio, Tech. Rep., 2022. [Online]. Available: https://cytrio.com/wp-content/uploads/2022/07/2022-Q2-State-of-CCPA-GDPR-Compliance-Research-Report-FINAL.pdf

[4] Deloitte, "The impact of automation on cybersecurity," 2023.