ComplianceGpt: An LLM-Assisted Regulatory Compliance Verifier

Priscilla Kyei Danso, Stony Brook University, New York, USA

I. Introduction

The widespread concern over mass data collection by social networks, corporations, and data brokers has prompted governments and jurisdictions to establish regulatory frameworks aimed at safeguarding personal data and enhancing privacy protections. Early frameworks like the Health Insurance Portability and Accountability Act (HIPAA) of 1996 introduced privacy and security standards for sensitive health information, including restrictions on unauthorized access to medical records and data confidentiality mandates. More recently, the European Union's General Data Protection Regulation (GDPR) introduced robust privacy rights such as data access, rectification, erasure, and portability, inspiring similar laws worldwide, including the California Consumer Privacy Act (CCPA).

However, adhering to these frameworks is challenging due to their complexity, jurisdictional variations, frequent updates, and the continuous need for audits and training. For individuals, understanding their rights under these intricate regulations can be equally overwhelming.

Large Language Models (LLMs) offer potential solutions by helping interpret and navigate complex legal texts, but they have significant limitations. LLMs are prone to inaccuracies and "hallucinations," producing unreliable or misleading responses that risk non-compliance. They also lack essential capabilities such as traceability, explainability, and the production of verifiable outputs, which are crucial in regulatory contexts. Even with fine-tuning, LLMs struggle with evolving regulations and the complex symbolic reasoning needed for compliance checking.

In contrast, formal verification approaches can effectively assess compliance by translating regulations into logical formulas using fragments of First-Order Logic (FOL) or First-Order Temporal Logic (FOTL). These methods evaluate whether disclosure events meet regulatory requirements by checking logical formulas against event logs. However, their main limitation is accessibility; they are not user-friendly for general audiences who must engage with formal logic languages to interact with these systems.

II. ComplianceGPT: A Hybrid Approach

To bridge this gap, we introduce ComplianceGPT, a hybrid system that combines specialized LLMs with a dedicated logic-based compliance checker, called précis. ComplianceGPT aims to provide an intuitive, user-friendly interface for navigating regulatory compliance while leveraging the rigor and reliability of formal verification. The system will process natural language queries (e.g., "Can a doctor send my medical record to a third party under HIPAA?") and translate them into FOL formulas (e.g., $disclose(p_1, p_2, q, \text{medical\_records}) \wedge \mathsf{inrole}(p_1, \text{doctor}) \wedge \mathsf{inrole}(p_2, \text{patient}) \wedge p_2 = q$, in which p_1 denotes the information sender, p_2 denotes the information receiver, and q denotes the subject whose PII is being released by p_1 to p_2) using a specialized vocabulary derived from the regulation's representation.

Realizing ComplianceGPT will involve addressing several technical challenges in both natural language processing (NLP) and formal verification fields. In NLP, we need to develop robust methods for translating natural language questions into precise logical representations compatible with formal compliance checkers. In formal verification, we must enhance compliance-checking algorithms to produce traceable and explainable outputs, such as compliance proof-trees that justify why a particular data disclosure is permissible under the applicable regulation.

A. ComplianceGPT Architecture

ComplianceGPT exposes a natural language input and output interface. The user's natural language query is processed by LLM1, which translates the query into a FOL representation understood by précis. Précis then evaluates this query against the relevant regulation, producing a compliance decision along with traceability information. This traceability information is subsequently processed by LLM2, which translates the results back into a user-friendly natural language response. Although the system appears to be a straightforward pipeline, its realization involves overcoming significant challenges:

- **Instantiating LLM1:** How do we develop or adapt an LLM that accurately translates user queries into the FOL representation required by précis?
- **Enhancing précis with Traceability:** How do we modify précis to generate detailed traceability information that explains its decisions?
- **Instantiating LLM2:** How do we ensure LLM2 effectively translates the traceability outputs from précis into concise, understandable natural language?

1) Instantiating LLM1: The most direct approach to instantiate LLM1 is to utilize a generic LLM (LLMG) capable of translating natural language into FOL. However, LLMG might produce formulas containing predicates outside the target regulation's vocabulary. To address this, LLMG can be fine-tuned or employed in an in-context learning configuration that restricts its output to the relevant predicates.

Training Data Generation: A critical step involves generating appropriate training data for LLM1. By leveraging the FOL representation of the regulation, we can systematically walk through the abstract syntax tree (AST) of the logic-based policy to create natural language training examples corresponding to each logical expression. For instance, given the FOL expression $disclose(p_1, p_2, q, \text{medical\_records}) \land \mathsf{inrole}(p_1, \text{doctor}) \land \mathsf{inrole}(p_2, \text{patient})$, we can create a variety of natural language queries such as "Can a doctor disclose a patient's medical records?" To diversify the training dataset, LLMG can be used to paraphrase these queries, generating multiple variations that improve the robustness of the fine-tuned model.

2) Enhancing Précis with Traceability Information: Extending précis involves augmenting its algorithms to provide detailed traceability and provenance information. This includes linking compliance decisions to specific sections of the regulation and generating execution trees that explain which rules were applied and why certain conditions were met or violated. Additionally, by incorporating symbolic execution techniques, précis can handle planning queries, such as determining conditions under which a doctor can share patient information under HIPAA.

3) Instantiating LLM2: Similar to LLM1, LLM2 translates the FOL-based outputs from précis into natural language. Initially, a deterministic translator can be used to generate rudimentary natural language explanations from the AST of the FOL representation. To refine these outputs, we can use LLMG with prompts designed to summarize and enhance the readability of the deterministic translations. Again, the paraphrasing trick can be employed to diversify and expand the training data, enabling either fine-tuning of LLMG or effective in-context learning for LLM2.

III. Evaluation Plan

1) **Baseline Comparison:** Our primary baseline involves a fine-tuned version of LLMG trained on the synthetic data generated using the techniques described above. We will compare ComplianceGPT's performance against this baseline across various evaluation metrics, including accuracy, completeness, and traceability of responses.
2) **Use of HIPAA Training Materials:** By comparing ComplianceGPT's responses to these standard training materials, we will identify any mismatches or discrepancies, which will help pinpoint weaknesses in our approach or highlight areas where ComplianceGPT can be further improved.
3) **Systematic Query Testing:** To further evaluate LLM1 and LLM2, we will generate random permutations of FOL queries using précis and test these in collaboration with the respective LLMs. For instance, outputs generated by précis can be translated back into natural language by LLM2 and fed into LLM1 to test whether it accurately reconstructs the original FOL query, and vice versa. This round-trip testing will help us evaluate the internal consistency of ComplianceGPT.
4) **Human-Subject Study:** In this study, fellow graduate students in the lab whose research focuses are computer security and/or AI/ML research will interact with both ComplianceGPT and the baseline system in a semi-adversarial setting, where their goal will be to elicit undesired or erroneous behaviors from the systems. This setting will help us assess the robustness of ComplianceGPT and gather qualitative feedback on user experience, system accuracy, and perceived trustworthiness.

IV. Ongoing Preliminary Work

To build a robust foundation for ComplianceGPT, we are currently engaged in research that explores the capabilities and limitations of LLMs in translating natural language (NL) into Propositional Linear Temporal Logic (LTL) formulas. This work is a critical step in understanding the effectiveness of LLMs in handling logical translations, a key component for the success of ComplianceGPT. Our investigations reveal that while LLMs are adept at generating basic logical formulas, they often falter when tasked with more complex translations that require deep contextual understanding and symbolic reasoning—limitations that directly inform the design of ComplianceGPT.

By thoroughly evaluating LLM performance on NL to LTL tasks, we gain insights into their shortcomings, such as inconsistency in logical accuracy and difficulty in adapting to new or evolving regulations. This foundational research informs the integration strategy for ComplianceGPT, ensuring that the LLM components are appropriately augmented with logic-based systems like précis to mitigate these shortcomings. Ultimately, the goal of ComplianceGPT is to provide an accurate, explainable, and adaptable tool for navigating regulatory compliance, leveraging the strengths of both AI-driven and formal verification methodologies.

V. Conclusion

ComplianceGPT aims to revolutionize the way organizations and individuals interact with regulatory frameworks by providing a reliable, traceable, and intuitive AI assistant. Our approach combines the strengths of LLMs in handling natural language with the rigor of formal methods, paving the way for more accessible and verifiable compliance solutions. While our current focus is on HIPAA, ComplianceGPT's modular design allows for easy adaptation to other regulations, offering a versatile tool for navigating the complex and evolving landscape of data privacy laws.