Evaluating the Compliance of Software Requirements to LGPD with LLM through Simulations

Lucas Moura[1,2], **Marcos Vinícius Lima**[1,2], **Allan Peixoto**[1,2], **Emanuel Coutinho**[1,2]

[1]Graduate Program in Computer Science (PCOMP)
[2]Federal University of Ceara – Quixadá – CE – Brazil
{lucasmoura07, mviniciuslp87, allanpeixoto}@alu.ufc.br
emanuel.coutinho@ufc.br

Abstract. *Protecting system user data is an increasingly relevant topic as legislation and regulations emerge to ensure security and privacy. This study presents a proposal for using LLM models to support the assessment of systems' compliance with the LGPD. This is an automated, requirements-based approach that enables simulations of various scenarios. Using LLMs to measure systems' compliance with the LGPD can result in benefits such as faster and more diverse solutions and reduced manual effort. However, risks such as analysis or interpretation errors, response quality, and dependence on accurate requirements specifications must be considered. This research is in an early stage, and the approach was evaluated on an ad hoc basis.*

1. Introduction

With the availability of systems for various purposes and the growing use of digital applications, the risk of potential data breaches may also increase. Furthermore, misuse or breach of user data may result in penalties. In this sense, designers and developers of such technologies must comply with regulations that guarantee user privacy, such as the General Data Protection Regulation (GPDR) [Parliament 2016], which in Brazil is covered by the LGPD [Brasil 2018] [Rocha et al. 2023]. In Brazil, software development organizations, whether public or private, that handle user personal data must comply with a large number of regulations and ensure that business and system requirements are legally compliant, i.e., they implement the LGPD in all their software systems [Canedo et al. 2020].

One approach to facilitating software development is the collaborative construction of Software Engineering models or artifacts [de Souza et al. 2012]. These elements serve as abstractions of the software being built, including requirements specifications, use case diagrams, class diagrams, and even the actual software source code. Models vary in their degree of formalism, presentation format, and documentation. This implies several possible scenarios and combinations.

In recent years, Large-Scale Language Models (LLM), such as GPT and Copilot, have emerged as relevant players in software development processes [Hou et al. 2024]. These models not only aid in code generation, but also influence architectures and workflows [Ozkaya 2023].

One way to adapt a system to the LGPD is to determine its level of compliance with the law. To do so, it is necessary to analyze aspects of the system in light of LGPD provisions. However, this is a demanding task. One way to measure a system's level of compliance with the LGPD is through checklists [Mendes et al. 2021], guides [Neves Camêlo and Alves 2023], questionnaires [Araújo et al. 2021], or interviews

[Neitzke et al. 2023]. In addition, it is necessary for someone familiar with the system to participate in clarification sessions, who can discuss subjective aspects of the law, and who can eventually adjust the system according to LGPD requirements and experiment with new scenarios. The objective of this research is to evaluate the compliance of systems with the LGPD with the support of LLM, using software requirements specifications for scenario simulations as a basis.

2. Conceptual Proposal and Automated Solution

To measure a system's compliance with the LGPD, it is necessary to have evaluation mechanisms, such as checklists [Mendes et al. 2021], guides [Neves Camêlo and Alves 2023], questionnaires [Araújo et al. 2021], or interviews [Neitzke et al. 2023]. However, this task is manual and subject to different interpretations, even considering the text of the law. One way to automate the evaluation and allow for experimentation with different scenarios is using LLM.

The purpose of this research is to create a mechanism that, based on prompts and requirements specifications, can analyze the system under study from the perspective of its compliance with the LGPD. Consequently, it can also identify risk areas and potential improvements. An example of a prompt to be used with the addition of a requirements specification as an attached file or in text would be:

> **Prompt**: Act as a senior consultant in LGPD compliance. Analyze the following Software Requirements Specification and provide an opinion on its adherence to the General Data Protection Law (Law 13.709/18). Follow the analysis structure below:
>
> 1. **Identification of Personal Data:** List and classify all personal mentioned data.
> 2. **Purpose and Legal Basis:** For each stated processing purpose, identify the corresponding proposed legal basis and assess its adequacy.
> 3. **Data Subject Rights:** Does the specification describe how the data subject's rights (Art. 18) will be met? Which ones?
> 4. **Security Measures:** Is there mention of security controls, retention periods, or anonymization?
> 5. **Sharing with Third Parties:** Does the document list operators/partners who will receive the data?
> 6. **Risk Analysis:** List the 3 main risks to compliance with the LGPD that you identify in the document.
> 7. **Executive Summary:** A summary with strengths, gaps, and recommendations.

With a pre-formatted initial solution, the possibilities expand: human review, solution refinement with feedback from the LLM, and generation of the final version of the report. Evaluation scenarios are simulated at various points in this process. If the solution reveals inconsistencies with any legal provisions, specific refinements can be made to provide a more appropriate response. If the solution returns indications of failures or the absence of some security aspect, for example, specific refinements can be made to generate a more complete response.

It is also possible to apply a questionnaire or checklist to a some specification. In general, the answers vary depending on the interpretation of the LLM, the level of detail in the specifications, and the questionnaire. This assessment is an alternative way to assess the level of compliance with the LGPD.

Based on a requirements specification, a prompt is defined for an LLM appropriate to the level of detail of the requirement information and output format. Another option

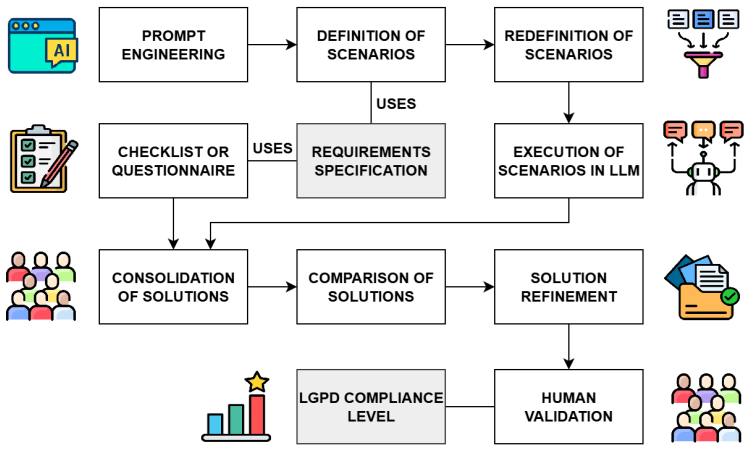

Figura 1. Proposed approach flow

is to use checklists or questionnaires that assess LGPD compliance based on the requirements specification. The scenario definition activity helps simulate compliance with the law and can be refined after refinements. Executing the prompts in the LLM yields results that, together with the questionnaire results, are consolidated and can be complementary. The solutions are analyzed and refined again, undergoing human validation at the end of the process. Finally, a level of LGPD compliance is determined for the system in question, which can be forwarded to a development team.

Figure 1 presents a flowchart of the proposed solution. This solution can be implemented through an application that integrates the user interface with a LLM. Additionally, the application's functionality can be extended to assess the quality of LLMs, qualitative user records, databases of potential legal flaws in requirements, software requirements aligned with legal articles, etc.

3. Preliminary Results

For the purpose of a preliminary evaluation of the approach, a simple example of the approach was used, following the evaluation flow of a functional requirements specification and a use case specification. Both were analyzed by an LLM, with a prompt similar to the one presented, and both were evaluated by a checklist proposed in [Mendes et al. 2021]. The subject of the documentation was a system for managing a university restaurant.

For the LLM analysis regarding the requirements specification, the following strengths were listed: provision for profile-based access control (authentication), which is the basis for security; and segregation of modules (Student, Professor, Administrator) demonstrating initial concern with the principle of need-to-know responsibilities. The gaps highlighted were: the system was designed for functionality, not for privacy. There are no requirements to meet the rights of data subjects (correction, deletion, portability); there is no legal justification for the processing of sensitive data, generally socioeconomic data; and requirements for retention, anonymization for reports, and transparency regarding sharing with the university system are lacking. And as recommendations: include a privacy module, with a dashboard where the data subject can download their data (portability), request corrections, and monitor the purpose of their data use; detail in the documentation that the socioeconomic form will have a specific consent step, informing

that refusal does not prevent the use of the rest of the system, but makes it impossible to apply for specific aid (consent linked to purpose); and implementation of an automatic routine for anonymizing or deleting application data, respecting the legal deadlines for storing public documents via log.

For the analysis of the LLM with the use case, the strengths were: concern with authentication and security for payments; and mention of communication failures and exception handling. For gaps: absence of provision for the rights of data subjects; lack of definition of retention and deletion periods; sharing with third parties without contractual details; and lack of mention of data anonymization. Finally, as recommendations: include in the system a module for addressing the rights of the data subject (Art. 18); define and document retention periods for each type of data; formalize a payment contract under the terms of Art. 39 of the LGPD; record access logs and ensure traceability; and revise the specification to include privacy principles from the design stage.

In both documents, aspects of Privacy by Design were highlighted, as neither was designed following these principles. The concept of Privacy by Design, proposed by Ann Cavoukian [Cavoukian 2010], addresses the incorporation of privacy and personal data protection from the very beginning of the product and service design process. There is a need for justifications to the legal bases and better data governance. For the use case specification, due to the format used, LLM returned that the system has an adequate functional structure, but requires significant adjustments to be in full compliance with the LGPD, especially regarding transparency, data subject rights, and third-party management.

Regarding the checklist on the specification of requirements, in general, although the functional requirements meet the business needs, the specification is silent regarding practically all non-functional requirements related to privacy and data protection. The system, as described, presents serious risks of non-compliance with the LGPD, as it handles data without providing for the mechanisms of transparency, security, and data subject rights required by law. Regarding the checklist on the specification of the use case, from the point of view of data protection, the specification is insufficient, as it treats personal data only as a means for the transaction, without considering the legal requirements of transparency, consent, security, and data subject rights.

As limitations and threats to the validity of this preliminary assessment, the human assessment was superficial, and may contain biases, misunderstanding of the checklist issue, different opinions in checking the issues, and superficial or non-detailed knowledge of the specifications. Regarding the use of LLM, it was basically the application of a prompt, without refinement, which may bring incorrect or misleading information from the assessment. Furthermore, only one LLM was used. Another aspect to consider is the security and privacy of the process itself, given the risk of sending potentially sensitive software requirements to third-party templates.

4. Challenges, Impacts and Relationship with Social, Human and Economic Aspects of Software Development

Several challenges and impacts can be listed as a result of applying the described approach. When they arise, the following situations may impact the development of software solutions, increasing or reducing costs, changing specifications, and increasing development effort. Challenges include: quality of assessment responses; lack of assessment

across all aspects of the LGPD; need for human validation; misinterpretation of the law; different LLMs offer different solutions and formats; increased effort for development teams; complexity of different simulation scenarios; and variety of methods for assessing LGPD compliance.

Regarding the technical, social, human, and economic aspects of software development, some considerations about the approach can be observed. Remember that this is a research proposal, which needs refinement and experimentation, so these are preliminary results. It is expected that after more elaborate experimentation with the approach, further considerations, improvements, and user feedback will be provided.

From a technical view, there are several situations. The agility in the involved tasks, such as refining and adjusting requirements, is a benefit. However, there is the issue of accuracy and reliability in the results, as LLM can provide incorrect or misleading information. In terms of productivity, there are scalable results, as the task execution time decreases, making it possible to increase the number of requirements. Regarding tools and practices, there is a need for adaptation or evolution, as there is a process change. Finally, the use of LLMs in requirements can provide false information, which raises ethical issues, topic closely related to the use of LLMs in everyday life.

Human and social aspects are directly related to research. The non-replacement of the professional is a situation that must be considered, as using an LLM often requires human validation. This may imply the need for adaptation of the requirements professional, who will use the LLM as a support tool. The feeling of eliminating the professional is something to be avoided, as there is a possibility that the professional will think that they will lose their usefulness or their job itself by using the LLM. However, there is a change in responsibilities, a need for professional adaptation. Finally, the impact on development teams can be beneficial, but it must be applied carefully.

Finally, from an economic aspect, there is the possibility of cost savings for projects and companies with the use of LLMs, however, other costs may be added. The need for professionals knowledgeable in LLMs, or at least who work well with Prompt Engineering, can be a cost. The impact of LLM results on requirements can be significant for projects, as poorly structured, erratic, or missing requirements may occur. Depending on the case, there may be a need for LLM licenses, and this may be an additional cost. Finally, the cost reduction/addition ratio should be evaluated by the team.

5. Final Remarks

This research proposes the use of LLMs to measure a system's level of compliance with the LGPD, based on requirements, and the possibility of analyzing different solutions through simulations of various scenarios. For a system or functionality, a preliminary or consultative analysis is necessary to establish criteria for evaluating the law. Subsequently, a checklist strategy can guide a process flow that will operationalize the development. Both steps can be generated by the presented proposal.

A suggestion for improving the research would be to investigate, cite, and adopt language models specifically trained or adjusted to Brazilian legislation, especially LGPD and its regulatory framework. Using a model with this profile that "knows" the LGPD, not just through prompt instruction, but because it has been trained with Brazilian legal

data, could bring significant benefits to the research. It can search for a model that has been trained with this data and cite it explicitly in the text.

Benefits include the possibility of obtaining faster and more diverse solutions regarding system compliance with the LGPD, and less human effort to generate responses. However, there are potential problems, such as incorrect assessments, specification dependence, and low solution reliability. The next steps include implementing the proposal and analyzing the impacts of its use.

Referências

Araújo, E., Vilela, J., Silva, C., and Alves, C. (2021). Are my business process models compliant with lgpd? the lgpd4bp method to evaluate and to model lgpd aware business processes. In *XVII Brazilian Symposium on Information Systems*, SBSI '21.

Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. lei geral de proteção de dados pessoais (lgpd).

Canedo, E. D., Calazans, A. T. S., Masson, E. T. S., Costa, P. H. T., and Lima, F. (2020). Perceptions of ict practitioners regarding software privacy. *Entropy*, 22(4).

Cavoukian, A. (2010). Privacy by design: The definitive workshop. a foreword by ann cavoukian, ph.d. *Identity in the Information Society*, 3(2):247–251.

de Souza, C. R. B., Marczak, S., and Prikladnicki, R. (2012). Desenvolvimento colaborativo de software. In Pimentel, M. and Fuks, H., editors, *Sistemas Colaborativos*, chapter 8. Elsevier, Rio de Janeiro.

Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J., and Wang, H. (2024). Large language models for software engineering: A systematic literature review. *ACM Transactions on Software Engineering and Methodology*, 33(8):1–79.

Mendes, J. a., Viana, D., and Rivero, L. (2021). Developing an inspection checklist for the adequacy assessment of software systems to quality attributes of the brazilian general data protection law: An initial proposal. In *Proceedings of the XXXV Brazilian Symposium on Software Engineering*, SBES '21, page 263–268.

Neitzke, C., Mendes, J. a., Rivero, L., Teixeira, M., and Viana, D. (2023). Enhancing lgpd compliance: Evaluating a checklist for lgpd quality attributes within a government office. In *XXII Brazilian Symposium on Software Quality*, SBQS '23, page 218–227.

Neves Camêlo, M. and Alves, C. (2023). G-priv: A guide to support lgpd compliant specification of privacy requirements. *iSys - Brazilian Journal of Information Systems*, 16(1):2:1 – 2.

Ozkaya, I. (2023). Application of large language models to software engineering tasks: Opportunities, risks, and implications. *IEEE Software*, 40(3):4–8.

Parliament, E. (2016). Général data protéction régulation (gdpr) - régulation (eu) 2016/679.

Rocha, L. D., Silva, G. R. S., and Dias Canedo, E. (2023). Privacy compliance in software development: A guide to implementing the lgpd principles. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, SAC '23, page 1352–1361.