Automated Public Transparency Auditing Using Web Scraping and Large Language Models for PNTP Compliance

Iasmim Maria Freire da Silva Torres[1], Cláudio de Souza Baptista[1], André Luiz Firmino Alves[2], Livia Aniely de Oliveira Almeida[1], Eniedson Fabiano Pereira da Silva Júnior[1]

[1]Federal University of Campina Grande (UFCG) – Campina Grande – PB – Brazil

[2]Federal Institute of Education of Paraíba (IFPB) – Picuí – PB – Brazil

{iasmim.maria.torres,livia.almeida}@ccc.ufcg.edu.br

baptista@computacao.ufcg.edu.br, andre.alves@ifpb.edu.br

eniedson.junior@copin.ufcg.edu.br

Abstract. *Public transparency, ensured by the Access to Information Law (Law No. 12.527/2011) and the National Public Transparency Program (Programa Nacional de Transparência Pública – PNTP), is still predominantly evaluated through manual processes, making it costly and difficult to scale. This study proposes an automated pipeline for auditing transparency portals of the Legislative Branch, combining web scraping and Large Language Models (LLMs) to assess compliance with PNTP criteria. Validation against human evaluations showed an average accuracy of 85.81%. The approach contributes to making transparency oversight more scalable, standardized, and traceable.*

1. Introduction

Public transparency is an essential condition for social control, governmental accountability, and oversight of public management. Access to information on budgets, expenditures, bidding processes, contracts, and other administrative acts allows citizens, regulatory agencies, and supervisory institutions to monitor the actions of public authorities. Empirical studies indicate that low levels of transparency are associated with a higher incidence of corruption and negative impacts on municipal health, education, and infrastructure indicators [Santos and Diniz 2024, Paiva et al. 2021].

In Brazil, Law No. 12,527/2011 (Access to Information Law – LAI) consolidated the duty of active and passive disclosure of public data. Within the scope of external control, the National Public Transparency Program (PNTP), coordinated by the Association of Members of the Courts of Accounts of Brazil (Atricon) in conjunction with the Courts of Accounts, structures the evaluation of transparency through a matrix composed of 176 criteria distributed across 26 thematic dimensions [ATRICON 2025].

Despite this normative reference, the verification of the conformity of transparency portals is still predominantly manual. Auditors need to individually examine each criterion of the matrix, evaluating the presence, quality, and manner in which information is made available. This process demands significant technical effort, is subject to interpretative variations, and presents scalability limitations, especially given the number of entities evaluated and the large number of criteria to be examined.

This scenario directly affects auditing institutions, Courts of Auditors, public oversight bodies, and technical teams responsible for monitoring transparency. In this context, web scraping techniques allow for the automation of information gathering on heterogeneous portals, while LLMs offer semantic interpretation capabilities for extensive and unstructured texts. The integration of these approaches presents itself as a promising alternative to support the systematic application of normative criteria on a large scale.

Given this, this work proposes and evaluates an automated pipeline for analyzing the Availability criterion in the dimensions applicable to the Legislative Branch, according to the PNTP matrix. The solution integrates web scraping and LLMs to extract, interpret, and classify information present in transparency portals, producing structured and justified outputs, amenable to human audit. The main contribution of the work lies in articulating automated data collection, semantic analysis guided by institutional criteria, and the generation of traceable evidence to support the work of auditing institutions and public regulatory agencies. The approach was validated through quantitative and qualitative experiments with real data from city councils in the state of Acre, in the year 2025, comparing the automated classifications with a template developed by experts.

The remainder of this paper is structured as follows: Section 2 presents related works; then, the methodology and architecture of the proposed solution are described on Section 3; subsequently, the results obtained are discussed on Section 4; and finally, Section 5 presents the conclusions and perspectives for future work.

2. Related Work

Literature on public transparency indicates that simply publishing information is not enough: citizens must be able to locate, understand, and monitor disclosed content. In this context, [Lourenço 2023] proposes a user-oriented transparency framework for public e-services, emphasizing traceability, metrics, and accountability from the portal design stage. Its main limitation is that the proposal remains mostly conceptual and does not define an automated and reproducible procedure for collecting evidence and verifying conformity with an institutional oversight framework.

Still focusing on access and usability, [Jácome Filho and Macêdo 2022] analyze municipal portals in mobile environments, showing that responsiveness and information architecture problems hinder navigation and content location. The study highlights how interface weaknesses reduce the social reach of transparency, especially where mobile phones are the main access channel. However, it does not propose continuous monitoring mechanisms or a way to transform usability findings into traceable and comparable evidence at scale.

With the increased use of automation in public services, transparency also depends on auditability and explainability. [Saldanha et al. 2022] evaluate Brazilian federal digital services and identify gaps related to criteria, risks, and justifications for data collection and use. Complementarily, [Kingsman et al. 2022] discuss the UK algorithmic transparency standard, proposing explanations for both the general public and technical audiences. Both studies reinforce the need for audience-appropriate justifications and audit support, but focus on general guidelines rather than on collecting evidence from heterogeneous portals and linking it to a specific oversight framework.

From an operational standpoint, portal heterogeneity and manual navigation li-

mit large-scale inspection. [Martins and Silva 2024] present a web scraping method for automating the download and collection of open data from transparency portals, demonstrating efficiency gains. However, the focus remains on extraction: automated collection alone does not ensure that evidence is interpreted, organized according to verifiable criteria, or clearly traced back to oversight requirements.

In compliance verification, [Iversen and Huang 2026] explore the use of large language models to interpret regulatory texts and support decisions in scenarios involving ambiguity and semantic dependencies. The work shows that language models can connect requirements to justifications, but tends to assume that relevant evidence is already available and well-formed. Systematically obtaining evidence from heterogeneous portals therefore remains a bottleneck for automated enforcement.

The use of artificial intelligence to strengthen public integrity has also been studied in anti-corruption initiatives. [Odilla 2023] analyzed 31 Brazilian anti-corruption bots, showing that many are developed by technically specialized public servants or citizens and face difficulties related to access to and standardization of open data. The study maps the potential and limits of automated monitoring, but does not propose a structured model of normative verification guided by a specific institutional framework.

The transparency debate also applies to AI systems used in the public sector. [Hsu and Anex-Ries 2025] propose a framework for evaluating governmental AI transparency, emphasizing explainability, governance, data documentation, and usage criteria from system design. The authors organize transparency into informational layers for citizens, experts, and auditors, reinforcing auditability and traceability. However, the focus is AI governance, not automated evidence collection in heterogeneous portals or its connection to external oversight criteria.

In summary, the reviewed studies show that effective transparency requires citizen-oriented design [Lourenço 2023], accessible interfaces [Jácome Filho and Macêdo 2022], explainability and auditability in automated systems [Saldanha et al. 2022, Kingsman et al. 2022, Hsu and Anex-Ries 2025], and computational techniques for large-scale collection and analysis [Martins and Silva 2024, Iversen and Huang 2026]. However, a gap remains in integrating automated data collection, normative interpretation guided by institutional criteria, and structured auditable evidence for expert validation.

Thus, this work proposes a pipeline that integrates web scraping and language models to evaluate the availability criterion in legislative branch portals according to the PNTP (National Policy for the Promotion of Public Administration). Unlike prior approaches, the solution combines automated extraction, criteria-guided interpretation, and structured auditable evidence generation.

3. Methodology

The methodology presented in this work describes the data processing flow used for the automated evaluation of public transparency portals in the context of the Legislative Branch.

The data analyzed corresponds to the transparency portals of the City Councils of the State of Acre, selected as the universe of the research because they present institutional

homogeneity and are subject to the same normative requirements within the scope of the PNTP (National Transparency Program). The geographical and institutional delimitation allowed for a comparative analysis between entities of the same administrative nature, reducing structural biases arising from differences between powers or federative spheres.

The data used in this study were obtained from the National Public Transparency Radar [1], a platform maintained by Atricon, considering the results related to the Availability dimension in the PNTP 2025 cycle. Thus, the data from Acre used in this study correspond to the consolidated results of the PNTP 2025 cycle, made publicly available on the Transparency Radar after the official announcement in December 2025. As institutional portals can be updated over time, the results analyzed reflect the situation recorded at the time of consolidation of the evaluation cycle, not necessarily including subsequent changes made by the evaluated agencies.

The analysis made in this work was focused on the availability of information, one of the five evaluation categories of the PNTP (National Transparency Program), which also includes timeliness, historical series, search filter, and recording of public information reports. This choice was motivated by the relevance of evaluating how transparency portals make information available, a crucial aspect to ensure the effectiveness of social control and oversight of public management. Furthermore, the scope was limited to the dimensions applicable to the Legislative Branch, specifically Dimensions 1 to 15 and Dimension 20 of the PNTP, in order to align the automated evaluation with the normative requirements pertinent to this type of jurisdiction and allow comparability between portals of the same institutional nature.

Regarding the model used, this work exclusively adopted GPT-5 in all inferences related to the evaluated dimensions. The adoption of a single model aimed to ensure methodological uniformity and guarantee greater consistency in the semantic interpretation of extensive, heterogeneous content with different presentation structures.

As for the model's usage configuration, the default parameters of the execution environment used were maintained, without manual adjustment of temperature, maximum token limit, or other generation hyperparameters. A specific strategy for dividing the content into smaller blocks was also not employed, since the texts extracted from the portals were submitted directly to the model within the available context limit. In cases where the content exceeded this limit, the input was truncated, and only the portion handled by the model was processed.

An automated abstention mechanism or explicit uncertainty handling was not implemented. Thus, the responses were analyzed based on the classifications produced by the model itself according to the criteria defined in the evaluation prompts.

The choice of the GPT-5 model is also due to its ability to handle lengthy and structurally varied texts, a particularly relevant characteristic considering the technological and organizational diversity of the portals analyzed.

3.1. System Architecture

The solution architecture was designed as a modular pipeline composed of three main stages: (i) extraction of portal content through web scraping with navigation simulation,

[1] `https://radardatransparencia.atricon.org.br/respostas.html`

allowing the complete loading of dynamic pages; (ii) preprocessing of HTML content, with removal of irrelevant elements and structuring of the text for analysis; and (iii) compliance analysis based on LLMs, in which the textual content is interpreted in light of the PNTP criteria applicable to the Legislative Branch. The inferences produced are stored in a structured format, enabling traceability, quantitative validation, and subsequent human auditing.

Figure 1 presents the overall system architecture, illustrating the processing flow from the data extraction stage to the generation of structured output for compliance analysis.

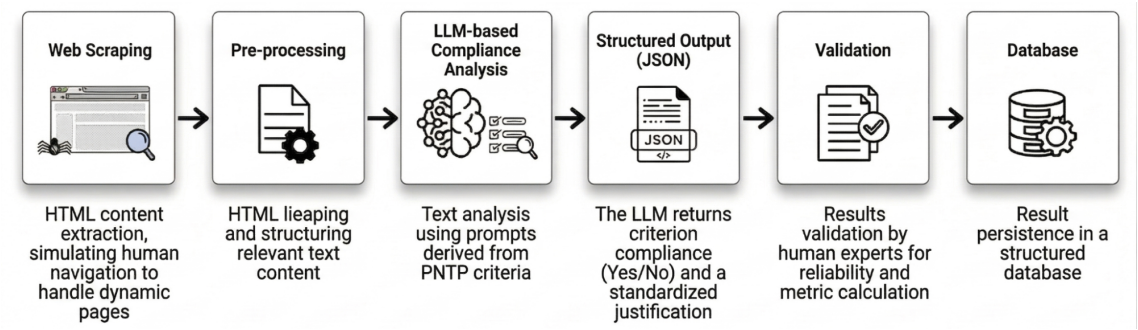

Figura 1. System Architecture

3.2. Database

At the beginning of the process, a database was structured containing the minimum information necessary to execute and track the automated evaluation of the portals. At the end of the data collection, 1,762 data records were obtained, corresponding to the units analyzed within the scope of this study. For each unit analyzed, the following were recorded: (i) the portal link (URL) to be collected; (ii) the name of the managing unit; (iii) the PNTP criterion associated with the verification; (iv) the evaluation item considered in this work (Availability); and (v) the final result of the analysis (Meets or Does Not Meet), subsequently consolidated after the inference and validation phase.

This organization ensured standardization of pipeline inputs, traceability of decisions, and support for consolidating results and calculating metrics.

3.3. Data Extraction and Preprocessing (Web Scraping)

The initial stage of the work consisted of extracting data from transparency portals, made possible by developing a pipeline using the Puppeteer[2] library in a Node.js environment. The choice of this tool was strategic to deal with the high technological heterogeneity of the pages, which vary between static HTML, dynamic content rendered via JavaScript, iframes, PDF documents, and images. To ensure the capture of information, the script was configured to operate in a two-level navigation model: in the first, the main page is accessed and the complete rendering of the DOM tree is ensured through controlled timeouts and instances of waiting for specific selectors.

In the second level, the collection flow performs a scan for iframe elements, performing context switching whenever the existence of iframes is detected, allowing data

[2]`https://pptr.dev/`

extraction. In a complementary way, the pipeline was programmed to identify and process images and PDF files, integrating the content of these documents into the analysis content. As a result of this stage, the complete HTML code of the analyzed pages was obtained, which underwent a cleaning and pre-processing process. This phase aimed to remove irrelevant elements, such as scripts and advertisements, as well as to structure the textual content necessary for the subsequent analysis stages.

3.4. Compliance Analysis with LLMs

At this stage, the pre-processed text content was submitted to the LLM-based compliance analysis module. For each criterion belonging to Dimensions 1 to 15 and 20 of the PNTP, the corresponding text was analyzed to verify compliance with normative requirements, exploring the model's ability to interpret semantics and context.

3.4.1. Prompt Engineering

The analysis was based on prompt engineering techniques, in which each criterion applicable to the Legislative Branch was translated into a structured instruction provided to GPT-5. The prompts explicitly describe the expected informational elements, guiding the model to identify evidence in the text indicating whether or not the criterion is met. In addition to the compliance decision, the model was instructed to present a textual justification based on the analyzed content, contributing to the interpretability of the process and allowing validation by human experts in light of the PNTP guidelines.

The input provided to the model consists of textual content extracted from the transparency portal page to be analyzed, including information potentially relevant to the evaluated criterion. This material is complemented by instructions derived from the defined persona, who assumes the role of the evaluator and establishes the normative context of the verification. Based on this set of information, composed of the extracted content and the normative guidelines, the model performs the conformity inference according to the PNTP criteria, producing the corresponding structured output.

Figure 2 presents the structure of a prompt used in this work, highlighting its main components, such as role definition (evaluator persona), contextualization, normative description of the criterion, analysis instructions, and decision generation with justification.

3.4.2. Structured Output

The LLM was configured to return its analyses in a structured JSON format, containing the assigned classification, the textual justification, and, when possible, the excerpt from the original text that supported the decision. This standardized structure allows for the systematic storage of inferences, as well as their subsequent traceability and validation by human auditors, in alignment with institutional control guidelines.

Beyond the aggregate metrics, the analysis of the structured outputs allows us to observe how the model justified its compliance decisions. In Figure 3, the model explicitly identified the consolidated values related to committed, liquidated, and paid expenses, correctly classifying the criterion as met. Figure 4, however, presents a case where the

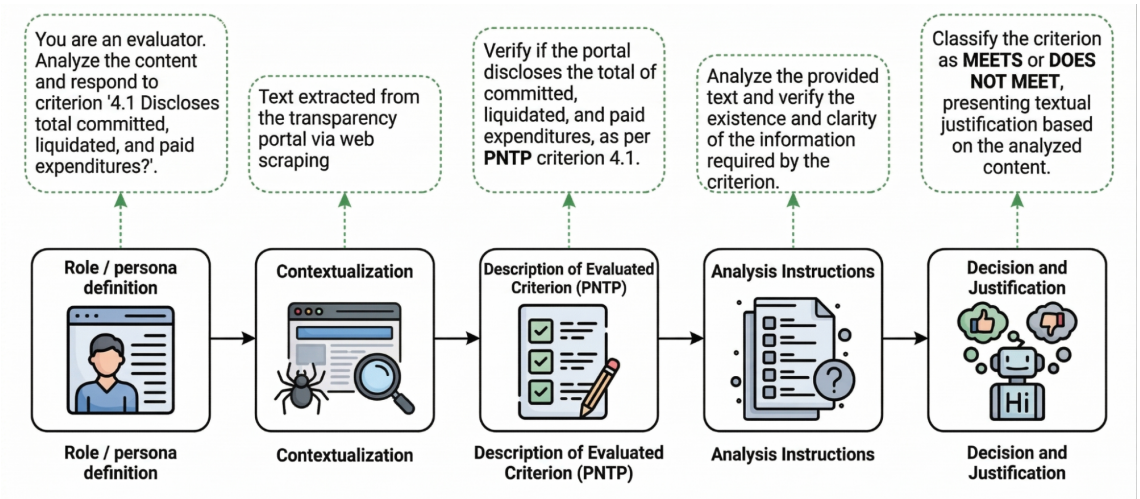

Figura 2. Logical structure of the prompt adopted in the analysis

criterion was classified as not met: although the portal offered data export functionalities, there was no explicit indication of the totals required by the normative criterion. These examples illustrate that the model is not limited to the generic presence of functionalities on the portal, but seeks to relate the extracted evidence to the informational elements specifically required by the PNTP.

```
{
    is0k: true,
    Justification: it was found in the following sections: \n
    "Totalizers - Commitment Value (Sum) R$ 13.789.261,83"; \n
    "Totalizers - Liquidated Value R$ (Sum) R$ 13.272.354,27"; n
    "Totalizers - Paid Value R$ (Sum) R$ 13.260.672,22"
}
```

Figura 3. Output of the model for compliance analysis - Meets

```
{
    "isOk": false,
    "Justification": "No. Justification for why it was not found:
        The content presents only export format options
        (PDF, ODT, ODS, CSV) and does not display values or fields that indicate
        the total committed, liquidated, and paid expenditures for the period,
        whether in a consolidated form or by summing the detailed information."
}
```

Figura 4. Output of the model for compliance analysis - Does Not Meets

The experts responsible for developing the template evaluated the portals independently, without prior access to the outputs produced by the model or necessarily to the same state of the portal accessed during the automated analysis. This strategy sought to preserve the autonomy of human evaluation, although it may have generated, in some cases, discrepancies resulting from changes in content, structure, or page availability

between analysis periods. Subsequently, the divergent cases were reviewed for correction and consolidation of the templates.

4. Results and discussion

The application of the proposed methodology resulted in the generation of automated classifications for the criteria of Dimensions 1 to 15 and 20 of the PNTP, in the context of the Availability assessment item. This section presents the quantitative and qualitative results arising from the comparison between the classifications produced by GPT-5 and the template developed by human experts.

4.1. Overall performance

Initially, the model's overall performance was evaluated using accuracy, precision, recall, and F1-score metrics. Across all criteria analyzed, the model showed an accuracy of 85.81%, precision of 95.72%, recall of 74.60%, and an F1-score of 83.85%. The comparison between the classifications produced by the GPT-5 and the human template resulted in 649 true positives, 863 true negatives, 29 false positives, and 221 false negatives.

The results demonstrate high overall performance, with accuracy substantially superior to recall. This behavior indicates that the model tends to adopt a more conservative stance in assigning the "Meets"class, reducing the risk of overestimating the compliance of the evaluated portals, but at the cost of failing to identify some of the criteria that are actually met. In the context of monitoring public transparency, this pattern can be considered desirable in scenarios supporting audits, as it prioritizes greater institutional reliability, even if it implies less sensitivity in some cases.

4.2. Performance by dimension and by criterion

Figure 5 presents the performance of GPT-5 in Dimensions 1 to 15 and 20 of the PNTP, considering the Availability evaluation item. A variation in performance is observed among the different normative axes. Dimensions 3 (100%), 4 (97.0%), 5 (97.0%), 13 (98.2%), and 15 (97.7%) showed the highest accuracy rates, while Dimension 10 registered the lowest performance (55.7%), followed by Dimensions 8 (70.8%), 2 (78.2%), and 6 (79.5%). This contrast highlights that the model's effectiveness varies according to the nature of the normative requirement being evaluated. In general, objective, textual, and explicitly presented verification criteria on the portal tend to produce better performance. Conversely, requirements that demand greater contextualization, consolidation of distributed data, or more in-depth technical interpretation present greater classification difficulties.

The breakdown by individual criteria reinforces this pattern. Among the best-performing criteria, all with a 100% success rate, the following stand out: 1.4 (existence of a content search tool on the portal), 1.3 (visibility of access to the transparency portal on the website's homepage), 12.2 (disclosure of the physical address, telephone number, email address, and operating hours of the SIC), 12.5 (disclosure of local regulatory instruments that govern Law No. 12,527/2011), and 11.2 (disclosure of the Management or Activities Report).

On the other hand, the worst-performing criteria were 6.4 (disclosure of the list of outsourced companies with detailed information, 40.91%), 2.9 (inclusion of the Public

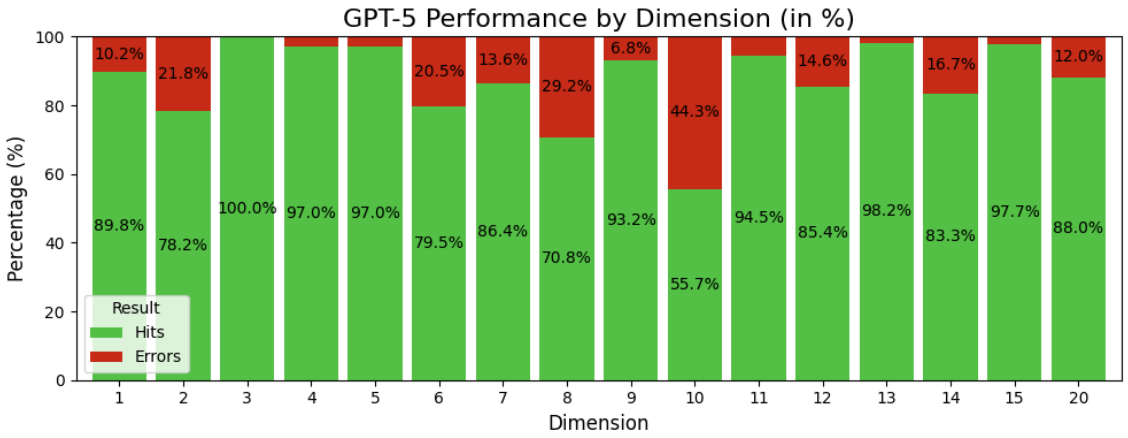

Figura 5. GPT errors x successes by dimension

Transparency Radar button, 36.36%), 8.1 (sequential disclosure of tenders with complete information, 36.36%), 8.3 (disclosure of the full text of documents from the internal and external phases of tenders, 36.36%), and 8.2 (disclosure of the full text of tender notices, 36.36%). These criteria generally involve lengthy documents, multiple procedural steps, distinct pages, varied formats, and distributed evidence, which increases the complexity of automated retrieval and interpretation.

Specifically regarding Dimension 10, related to public works, it was found that many portals provided information in external spreadsheets, often hosted on services like Google Sheets, sometimes with authentication or permission restrictions. In these situations, the automated extraction step was unable to fully retrieve the data, causing the model to operate on incomplete or inaccessible content. This factor directly contributed to the increase in false negatives and the lower performance observed in this dimension.

4.3. Main errors, limitations and implications

The observed errors can be grouped into three main categories: limitations in data extraction, semantic and normative ambiguities, and inherent restrictions in the automated interpretation process performed by LLMs. The heterogeneity of the portals, with dynamic content, PDFs, tables, and information distributed across multiple pages, hindered the complete collection of evidence. Furthermore, in several cases, the portals offered search or export mechanisms but did not explicitly present the elements required by the PNTP, which contributed to false negatives. Finally, criteria involving clarity, completeness, and adequacy of presentation require more contextual interpretation, making automated recognition of compliance more difficult when evidence is implicit or fragmented.

In addition to these operational limitations, the use of LLMs in the context of public auditing requires attention to the risks of hallucination and algorithmic bias. Because language models produce responses based on probabilistic patterns, there is a possibility of generating justifications that are not fully supported by the evidence extracted from the portals, as well as excessively flexible or restrictive interpretations of the evaluated criteria. This risk is especially sensitive in public transparency assessments, as erroneous classifications can generate false positives, by indicating compliance without sufficient evidence, or false negatives, by failing to recognize valid information present on the por-

tal.

It is also observed that the model's performance can be influenced by how each portal organizes and presents its information. Portals with a more standardized textual structure, clearer navigation, and explicitly described content tend to be more easily interpreted by the model, while portals with fragmented information, non-uniform language, PDF documents, or predominantly visual elements may increase the chance of error. Thus, the results produced by the system should be understood as support for analysis, and not as substitutes for human decision-making in auditing processes. To mitigate these risks, the approach adopted utilized structured prompts and explicit evaluation criteria, in addition to comparison with human reference classifications. Even so, the results indicate the need for additional mechanisms in future applications, such as recording the evidence used by the model, human review of ambiguous cases, abstention strategies in the face of insufficient information, and the use of evidence retrieval techniques to better support the generated responses.

4.3.1. Illustrative examples

To complement the analysis of the observed errors, two illustrative examples of divergence between the classification produced by the model and the human assessment are presented. In a case of a false negative in Figure 6, related to criterion 6.3 ("Does it disclose the list of its interns?"), the model classified the portal as "Does Not Meet," justifying that the page contained only images and an external link, without explicit text with the list of interns, start and end dates of contracts, or a formal declaration of non-existence. However, the human assessment considered that the portal met the criterion, as it presented an explicit justification for the absence of the required data. In this case, the error resulted from the model's inability to recognize that the formally justified absence of the information also constituted compliance with the normative requirement, resulting in an undue negative classification.

```
{
    isOk: false,
    Justification: "No. Reason why it was not found: The page contains only images and an external link,
    without any text presenting a list of interns with the start and end dates of their contracts, nor an
    explicit and dated statement of non-existence indicating the specific assignments."
}
```

Figura 6. Example of false negative

In contrast, a false positive is also observed in Figure 7, related to criterion 15.5 ("Regulates Federal Law No. 14,129/2021 (Digital Government) and discloses the regulations on its portal?"), in which the model classified the portal as "Complies" based on excerpts from Law No. 14,129/2021 and general references to the principles of digital government. However, the evaluated criterion required the disclosure of specific regulations from the agency itself, and not just mention of general legislation applicable to public administration. In this case, the model misinterpreted the regulatory requirement, improperly associating the existence of a generic legal basis with compliance with the specific criterion of the PNTP. This type of error highlights a semantic limitation of the inference, in which the model identifies content related to the topic but does not adequately dis-

tinguish the level of specificity required by the question, resulting in an undue positive classification.

```
{
    isOk: true,
    Justification: "Yes. It was found in the following excerpts: 'LAW No. 14,129, OF MARCH 29, 2021 ...
    Provides for principles, rules and instruments for Digital Government and for increasing public
    efficiency...'; 'CHAPTER II ON THE DIGITALIZATION OF PUBLIC ADMINISTRATION AND THE DIGITAL PROVISION
    OF PUBLIC SERVICES - DIGITAL GOVERNMENT'; 'Article 3. The principles and guidelines of Digital
    Government are...'"
}
```

Figura 7. Example of false positive

In general, the findings indicate that integrating web scraping and Large Language Models is technically feasible for supporting PNTP compliance analysis, especially in initial screening, evidence organization, and prioritization of cases requiring human review. In institutional auditing contexts, the system can increase scalability and traceability by reducing repetitive work, but it should not replace auditors. Its adoption requires caution because performance varies by criterion: objective and explicit requirements tend to be classified more reliably, whereas criteria involving deep navigation, distributed evidence, external documents, or contextual interpretation remain more challenging. Thus, the tool should be used under human supervision, with periodic error analysis and continuous calibration to mitigate technical bias and the under-identification of existing conformities.

5. Conclusions

This work presented an approach for the automated evaluation of public transparency portals through the integration of web scraping techniques and LLMs, applied to the PNTP criteria matrix. The methodology was evaluated by comparing the classifications produced by the GPT-5 model with a template previously developed by human experts.

Empirical results demonstrated that the proposed methodology is capable of reproducing, with a high degree of alignment, the classifications performed by human experts, achieving an accuracy of 85.81% and an F1-score of 83.85%. Analysis by dimension and by criterion indicated that the model's performance varies according to the nature of the regulatory requirement being evaluated, showing greater adherence to objective verification criteria and lower performance in requirements that demand contextual interpretation or consolidation of distributed information.

As a contribution, the study demonstrates the technical feasibility of using LLMs as a tool to support large-scale regulatory oversight, integrating automated extraction, semantic interpretation, and the generation of structured evidence. Future developments include comparing different language models, improving the evidence extraction and retrieval stages, refining the prompts used in compliance analysis, and incorporating Retrieval-Augmented Generation (RAG) techniques. These strategies can help reduce the occurrence of false negatives, especially given the 74.60% recall rate, which indicates that some of the existing compliances were not recognized by the system.

Also highlighted as future work is the improvement of the handling of external evidence used by the portals, such as spreadsheets, shared documents, and links to third-party services. The analysis of dimension 10, relating to public works, indicated that

external resources with restricted access, such as non-public Google Sheets spreadsheets, can limit or interrupt the extraction pipeline and compromise the automated verification of compliance. Future versions of the approach should differentiate between a lack of information and technical access failures, explicitly record these occurrences, and incorporate alternative data collection or human review strategies for cases where evidence is unavailable due to permission restrictions.

Finally, it is proposed to expand the evaluation to other categories of the PNTP and extend the methodology to portals of the Executive and Judicial branches, allowing for the assessment of its generalizability in different institutional contexts.

Referências

ATRICON (2025). Cartilha do programa nacional de transparência pública.

Hsu, E. and Anex-Ries, Q. (2025). A framework for assessing ai transparency in the public sector.

Iversen, O. and Huang, L. (2026). Leveraging large language models for bim-based automated compliance checking. *Automation in Construction*, 182:106707.

Jácome Filho, E. A. and Macêdo, J. M. A. (2022). Analysis of responsiveness and usability in websites serving public transparency in a mobile environment: Case study in the state of paraíba through heuristic evaluation. In *International Conference on Human-Computer Interaction*, Cham. Springer International Publishing.

Kingsman, N. et al. (2022). Public sector ai transparency standard: Uk government seeks to lead by example. *Discover Artificial Intelligence*, 2(1).

Lourenço, R. P. (2023). A framework for public eservices transparency. *International Journal of Electronic Government Research (IJEGR)*, 19(1):1–19.

Martins, C. P. P. and Silva, G. S. (2024). Método e aplicação de automatização para download de dados abertos em portais de transparência. In *Simpósio Brasileiro de Sistemas de Informação (SBSI)*. Sociedade Brasileira de Computação (SBC).

Odilla, F. (2023). Bots against corruption: Exploring the benefits and limitations of ai-based anti-corruption technology. *Crime, Law and Social Change*, 80(4):353–396.

Paiva, M. E. R., Ribeiro, L. L., and Gomes, J. W. F. (2021). O tamanho do governo aumenta a corrupção? uma análise para os municípios brasileiros. *Revista de Administração Pública*, 55(2):272–291.

Saldanha, D. M. F., Dias, C. N., and Guillaumon, S. (2022). Transparency and accountability in digital public services: Learning from the brazilian cases. *Government Information Quarterly*, 39(2):101680.

Santos, M. S. X. and Diniz, M. B. (2024). Determinantes da corrupção municipal no setor educacional brasileiro entre os anos de 2011 e 2014. *Semestre Econômico*, 27(62).