Using ChatGPT to Support Accessibility Requirements Specification: Experience from a Public Sector Platform

Anna B. Marques[1][0000−0001−9214−3399], Rhenara Oliveira[2][0009−0002−2501−9800], Ramon Luna[2][0009−0002−8892−9369], Amanda Sousa[2][0000−0001−6379−2648], José Antônio Macêdo[2][0000−0002−0661−2978], and Rossana Andrade[2][0000−0002−0186−2994]

[1] Federal University of Ceará (UFC), Russas, Ceará, Brazil
[2] Federal University of Ceará (UFC), Fortaleza, Ceará, Brazil
{beatriz.marques,rossana}@ufc.br,
{rhenaraalves21,ramonnicolas}@alu.ufc.br, amandasousa@great.ufc.br,
jose.macedo@insightlab.ufc.br

Abstract. Accessibility is a key quality requirement for digital systems, especially in the public sector, where inclusive access is essential. However, specifying accessibility requirements remains challenging in Requirements Engineering, as it requires specialized knowledge and careful interpretation of standards such as WCAG. Although Large Language Models (LLMs) have been explored for requirements-related tasks, there is limited empirical evidence of their use for accessibility requirements in real-world projects. This paper reports an industry experience using ChatGPT to support the specification of accessibility requirements for the Plataforma Antonieta de Barros, a public-sector digital platform designed to support data-driven decision-making in education. The goal is to investigate how an LLM can assist in generating requirements aligned with WCAG 2.2 and to evaluate their quality from practitioners' perspectives. The findings suggest that ChatGPT can support the initial specification of clear, relevant accessibility requirements, but human validation and iterative refinement are essential to ensure quality and applicability. The experience also highlights the importance of structured prompts and multidisciplinary evaluation to transform AI-generated outputs into actionable requirements for real-world software projects.

Keywords: Accessibility Requirements · Large Language Models (LLMs) · Requirements Engineering.

1 Introduction

Digital accessibility is a fundamental quality requirement for contemporary software systems, particularly in the public sector, where services must be inclusive and available to a diverse population [10]. However, accessibility is still insufficiently addressed during software development, especially in early Requirements Engineering (RE) activities [8, 9].

One key challenge is specifying accessibility requirements, commonly treated as Non-Functional Requirements (NFRs). Eliciting and documenting NFRs is difficult because stakeholders may struggle to express quality concerns, while software teams may lack domain knowledge or practical guidance. In practice, practitioners still face difficulties in identifying, formalizing, and prioritizing accessibility requirements [8, 9].

The rise of Large Language Models (LLMs), such as ChatGPT, has created new opportunities to support RE tasks. Prior work has explored LLMs for general RE activities [7], eliciting knowledge for NFR catalogs [5], and identifying domain-specific quality attributes in early elicitation stages [4]. Sampaio et al. [6] found substantial overlap between AI-generated and human-generated requirements, while highlighting the need for supervision. Arora et al. [1] discussed how LLMs may support elicitation, specification, analysis, and validation. Brotosaputro et al. [2] showed that AI-based solutions can help identify and mitigate digital accessibility barriers. These studies suggest that LLMs can help practitioners access domain knowledge, identify relevant concerns, and accelerate requirement-related activities. However, there is still limited empirical evidence regarding the use of LLMs specifically for accessibility requirements specification in real-world projects.

To address this gap, this paper reports an industry experience using ChatGPT to support the specification of accessibility requirements during the development of the *Plataforma Antonieta de Barros*, a public-sector platform designed to support data-driven decision-making in education.

The objective of this paper is to investigate how an LLM can support the generation and refinement of accessibility requirements aligned with WCAG 2.2 Level A, and to evaluate the perceived quality and usefulness of the resulting specification from the perspective of project practitioners.

2 Case Description

This case study was conducted during the development of the *Plataforma Antonieta de Barros*, a public-sector digital solution designed to support governance and data-driven decision-making in education. The platform was developed through a partnership between the National Fund for Educational Development (FNDE) and the Federal University of Ceará (UFC), with the goal of supporting the design, monitoring, and evaluation of educational public policies.

The platform was conceived as a data ecosystem that integrates multiple information sources related to FNDE programs and initiatives. It uses data science and AI techniques to support large-scale data analysis, evidence-based decision-making, transparency, and a data-driven culture within FNDE.

Its target audience is diverse, including managers, analysts, researchers, and citizens, which implies different levels of digital literacy and heterogeneous accessibility needs. Given its public nature and broad audience, accessibility was treated as an essential quality requirement. The project adopted the Web Content Accessibility Guidelines (WCAG) 2.2 Level A [10], organizing requirements according to the four principles: Perceivable, Operable, Understandable, and Robust.

The team also identified the need to document accessibility requirements in a structured, operational manner. Prior literature shows that eliciting non-functional requirements, especially accessibility-related ones, remains challenging due to difficulties in stakeholder communication and the need for specialized knowledge [3].

In this scenario, ChatGPT was used to support the specification of WCAG-aligned accessibility requirements in a domain that demands specialized expertise. Recent studies have highlighted the potential of LLMs to support RE activities, including elicitation and specification of non-functional requirements [1, 8, 9].

The initial expectation was that the model could expand requirement coverage and improve specification consistency. However, human analysis and validation were considered essential throughout the process. As a result, the project produced accessibility requirements organized according to the WCAG principles, covering aspects such as text alternatives, keyboard navigation, error identification, and assistive technology compatibility.

3 Solution and Results

3.1 Accessibility Requirements Generation Process

The generation of accessibility requirements with ChatGPT was conducted in five stages: prompt experimentation, output comparison, and consolidation of the resulting specification.

Stage 1 – Initial tests. Different prompts were tested using free and paid versions of ChatGPT. Both versions generated relevant requirements, but the paid version produced more detailed and context-aware outputs. Some responses exceeded the requested scope, reinforcing the need for prompt refinement.

Stage 2 – Similarity analysis. Outputs from different executions were compared to identify convergences, redundancies, and gaps. Recurrent requirements involved text alternatives, keyboard navigation, link identification, and compatibility with assistive technologies.

Stage 3 – Final prompting strategy. Based on prior results, the team adopted the paid version with a contextualized prompt requesting accessibility requirements aligned with WCAG 2.2 Level A for each user story.

> *I am developing a software system and would like your help specifying the accessibility requirements. The requirements must comply with WCAG 2.2, Level A. Please consider the context of the Antonieta de Barros Platform and the following feature description. Given this context, act as a requirements engineer or software development professional seeking to identify accessibility requirements. Specify the accessibility requirements applicable to the <user story specification> functionality.*

Stage 4 – Traceability matrix. A traceability matrix was created to compare requirements across executions and user stories, helping identify equivalent, complementary, or redundant items. Full coverage was observed for Perceivable and Robust principles, while Operable requirements varied depending on functionality.

Stage 5 – Final consolidation. Requirements were merged into a single specification, removing redundancies, standardizing wording, and organizing items according to the WCAG principles: Perceivable, Operable, Understandable, and Robust. Table 1 provides an overview of the types of accessibility requirements generated, organized by WCAG 2.2 principles.

Table 1. Summary of accessibility requirements organized by principle

Principle	**Types of requirements specified**
Perceivable (4)	Text alternatives for icons and images; semantic identification of fields and groupings; appropriate use of color; minimum contrast for text and visual elements
Operable (4)	Full keyboard navigation; logical tab order; clear definition of the purpose of links and buttons; descriptive page titles; absence of navigation barriers (e.g., controllable modals)
Understandable (4)	Definition of page language; control of context changes on focus; clear error messages associated with fields; explicit labels and instructions for data entry
Robust (1)	Appropriate use of semantic attributes and ARIA; compatibility with assistive technologies; correct updating of states in dynamic components

The final specification included 13 requirements: Perceivable (4), Operable (4), Understandable (4), and Robust (1). Robust was less represented due to implementation-oriented concerns, such as semantic markup, ARIA, and assistive technology compatibility. The public repository[3] contains the specification, prompts, evaluation instruments, anonymized results, refinement notes, and comparison artifacts.

An example requirement was:

> **Error Identification (Understandable):** Error messages must appear near the affected field and clearly describe the problem (e.g., “Name is required”). Screen readers must correctly identify and read these messages.

3.2 Evaluation of Generated Requirements

Participants The first evaluation involved 12 members of the *Plataforma Antonieta de Barros* team acting only as evaluators who had not participated in

[3] https://zenodo.org/records/19827189

the original specification of these requirements. Participants represented multiple areas, including development (5), UX/UI (3), requirements (2), testing (2), and research (1), with one participant indicating multiple roles. This profile enabled the assessment of the specification from different perspectives, ranging from implementation concerns to usability and quality issues.

Evaluation Instrument A questionnaire was used to assess the requirements organized according to the four WCAG 2.2 principles: Perceivable, Operable, Understandable, and Robust. For each category, participants evaluated five criteria: relevance, pertinence, innovation, clarity, and sufficiency. A five-point Likert scale was adopted. These criteria were inspired by Sampaio et al. [6], who evaluated the use of LLMs in Requirements Engineering, and were adapted to the context of accessibility requirements. The instrument also included open-ended questions to collect suggestions for improving the specification.

Results Figure 1 presents the distribution of responses. Overall, results indicate a strongly positive perception of the generated and consolidated requirements. In all categories, most responses were concentrated in *Agree* and *Strongly Agree*, with few neutral or negative responses.

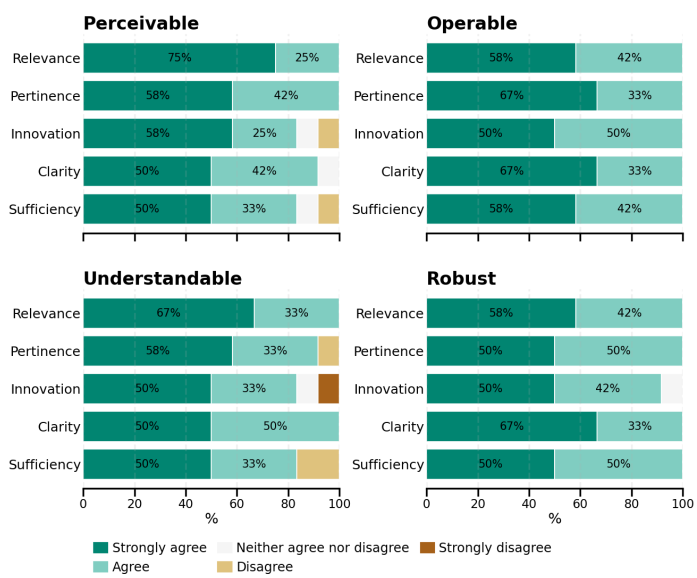

Fig. 1. First evaluation of accessibility requirements by WCAG category.

Qualitative comments complemented these findings. For the **Perceivable** principle, participants requested greater detail on color contrast and visual hierarchy. For **Understandable**, respondents suggested more practical examples and clearer organization of information. For **Robust**, comments highlighted the need for practical support, such as workshops on accessibility testing and validation with assistive technologies. Overall, the first evaluation suggests that the generated requirements were considered relevant and useful. The main opportunities for improvement involved additional technical detail, concrete examples, and stronger implementation guidance.

3.3 Requirements Refinement

Based on participant feedback, three main improvement needs were identified: (i) greater technical detail, (ii) inclusion of practical examples, and (iii) better clarity and standardization of wording. Therefore, the prompt was redesigned to request a more structured specification model that includes: requirement ID, guideline name, responsible area (UX/UI or Development), user impact, a contextualized example, and verifiable criteria.

This refinement produced requirements that were clearer, more actionable, and more implementation-oriented. The new structure also helped bridge communication between technical and design teams.

An example of a refined requirement is shown below:

R11 – Error Identification
Reference: UX/UI and Development
User Impact: Vague messages make problem resolution difficult.
Example: A user attempts to restore an item that has already been deleted. If the system only disables the button without explanation, the reason is unclear.

- Error messages must appear near the affected field and clearly describe the issue.
- Screen readers must correctly identify and read associated messages.

3.4 Second Evaluation of Refined Requirements

After refinement, the revised requirements were evaluated by development team members who had not participated in the original specification. The goal was to assess comprehension and the feasibility of implementation. Participants evaluated five criteria: relevance, clarity, sufficiency of implementation, feasibility, and ease of understanding. As shown in Figure 2, responses again concentrated in positive categories, indicating improved perceptions after refinement.

Relevance achieved near-unanimous agreement across all WCAG principles. Clarity was also highly rated, particularly for the Operable category, suggesting that the revised structure made requirements easier to understand. Implementation sufficiency showed greater dispersion, especially for Understandable and Robust requirements, indicating that some participants still expected more technical detail or practical examples. Feasibility remained positive overall, although keyboard navigation and complex interactive components were perceived as more challenging. Ease of understanding presented slightly greater variability, particularly for semantic markup, ARIA attributes, and more abstract behaviors.

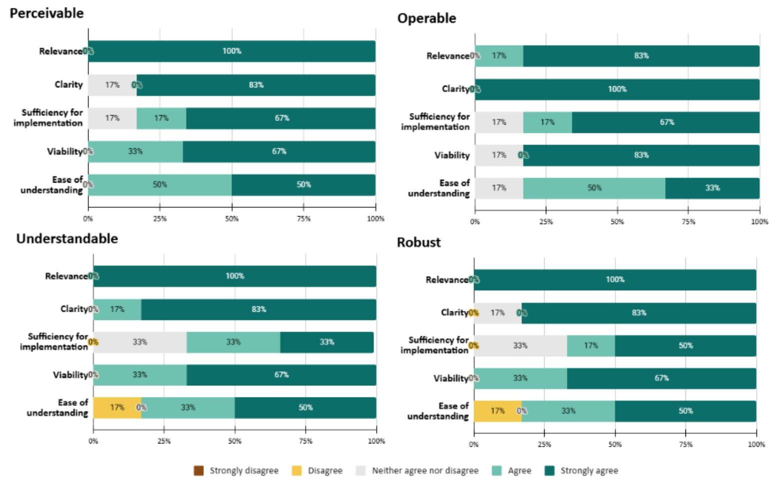

Fig. 2. Second evaluation of refined accessibility requirements.

Qualitative comments highlighted WCAG alignment, concrete examples, and next steps such as prioritizing requirements, defining validation criteria, and strengthening collaboration between development and design teams. Overall, refinement improved clarity, usefulness, and implementation readiness, while revealing the need for support materials for more complex requirements.

4 Conclusion and Lessons Learned

This paper reports an industry experience using ChatGPT to support accessibility requirements specification for a public-sector platform. ChatGPT supported ideation, requirement generation, and WCAG coverage, but human review was essential to contextualize outputs, remove redundancies, and make requirements actionable. Abstract topics, such as keyboard interaction, semantic markup, and ARIA-related behaviors, still required stronger technical guidance.

Lessons learned include: (L1) Contextualized prompts produced substantially better results than generic prompts; (L2) Human validation is essential to ensure consistency, feasibility, and alignment with project needs; (L3) Structured output templates improved the clarity and usefulness of generated requirements; (L4) Multidisciplinary evaluation involving development, UX/UI, testing, and requirements teams enriched the refinement process; (L5) Some accessibility requirements still require complementary implementation guidelines beyond textual specification alone.

Although limited to one project, one LLM, and without comparison with manual specifications, the findings provide practical, context-dependent evidence of how generative AI can support the generation of accessibility requirements. Future work includes replication in other domains and comparison with accessibility specialists.

Acknowledgements

We thank FNDE for funding the project "Big Data and AI Platform for the Governance of Basic Education Programs for the FNDE" (12.222/2023), as well as FUNCAP, INCT INES IA, and CNPq for supporting this work, including the productivity grant awarded to Rossana M. C. Andrade (306362/2021-0).

References

1. Arora, C., Grundy, J., Abdelrazek, M.: Advancing requirements engineering through generative ai: Assessing the role of llms (2023), https://arxiv.org/abs/2310.13976
2. Brotosaputro, G., Supriyadi, A., Jones, M.: Ai-powered assistive technologies for improved accessibility. International Transactions on Artificial Intelligence (ITALIC) **3**, 76–84 (11 2024). https://doi.org/10.33050/italic.v3i1.645
3. Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J.: Non-Functional Requirements in Software Engineering. Springer (2000). https://doi.org/10.1007/978-1-4615-5269-4
4. Portugal, R.L.Q., Silva, L.F., Sousa, H.P., Almentero, E.: Utilizing llms for early-stage qualities elicitation: A study on responsible fact-checking in journalism. In: Anais do Workshop em Engenharia de Requisitos - Proceedings of the 28th Workshop on Requirements Engineering (WER2025) (2025). https://doi.org/10.29327/1588952.28-29
5. Portugal, R.L.Q., Silva, L.F., Sousa, H.S.P., Leite, J.C.S.P.: Exploring llms on supporting the elicitation of knowledge for nfr catalogs: Insights from the transparency case. In: Anais do Workshop em Engenharia de Requisitos - Proceedings of the 28th Workshop on Requirements Engineering (WER2025) (2025). https://doi.org/10.29327/1588952.28-11
6. Sampaio, S.S., Lima, M.S., de Souza, E.R., Meireles, M.A., Pessoa, M.S., Conte, T.U.: Exploring the use of large language models in requirements engineering education: An experience report with chatgpt 3.5. In: Proceedings of the XXIII Brazilian Symposium on Software Quality. p. 624–634. SBQS '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3701625.3701687
7. Silva, Y., Gois, M., Santos, A., Castro, J., Lencastre, M.: Uso de large language models na engenharia de requisitos. In: Anais do Workshop em Engenharia de Requisitos - Proceedings of the 28th Workshop on Requirements Engineering (WER2025) (2025). https://doi.org/10.29327/1588952.28-18
8. Vinadé, R., Marczak, S.: Unveiling developer perspectives: A survey on accessibility practices and requirements in software development. In: Anais do Workshop em Engenharia de Requisitos - Proceedings of the 27th Workshop on Requirements Engineering (WER2024) (2024). https://doi.org/10.29327/1407529.27-15
9. Vinadé, R., Goelzer, N., Marczak, S.: On the definition of a method for specifying accessibility requirements: Early insights from a focus group with practitioners'. In: Anais do XXXIX Simpósio Brasileiro de Engenharia de Software. pp. 769–775. SBC, Porto Alegre, RS, Brasil (2025). https://doi.org/10.5753/sbes.2025.11561
10. W3C: Web content accessibility guidelines (wcag) 2.2. https://www.w3.org/TR/WCAG22/ (2023), accessed: 2026-04-05