

Generating Inclusive Personas with Documentary Evidence: An Exploratory Study Using LLMs and RAG

Ana Gabriela Pires e Silva
Universidade Federal do Amazonas
Instituto de Computação
Manaus, Amazonas, Brazil
gabriela.pires@ufam.edu.br

Tayana Uchôa Conte
Universidade Federal do Amazonas
Instituto de Computação
Manaus, Amazonas, Brazil
tayana@icomput.ufam.edu.br

Abstract

Inclusive personas are design artifacts that support design and requirements engineering by representing the needs, barriers, and contexts of marginalized users without reducing them to homogeneous or stereotyped profiles. However, creating such personas is challenging, since they must be useful for design decisions while also preserving contextual specificity and representational care. Large Language Models (LLMs) offer a scalable way to generate personas, but relying only on the model's general knowledge may lead to generic, weakly grounded, or stereotyped representations. This paper presents a preliminary approach for generating inclusive personas with LLMs supported by a document-based dataset and Retrieval-Augmented Generation (RAG). Rather than using RAG as a generic retrieval mechanism, the approach transforms documentary evidence about LGBTQIA+ people's experiences into categorized, metadata-enriched records, organizes them in a RAG-ready retrieval index, and uses them to ground the representation of needs, barriers, support networks, service interactions, and design opportunities in the generated personas. We evaluate the approach through an exploratory study with four evaluators, comparing personas generated with and without dataset support across three usage scenarios. The personas were analyzed from two complementary perspectives: technical/structural quality and diversity. The results suggest that documentary grounding can improve specificity and sensitivity to diversity in some scenarios, but it does not automatically guarantee better personas. When retrieved evidence is poorly aligned or thematically unbalanced, grounding may also introduce distortions or excessive emphasis on sensitive themes. These findings highlight both the potential and the limitations of using document-based grounding to support inclusive persona generation.

Keywords

Personas, Large Language Models, Retrieval-Augmented Generation, Inclusive Design, Requirements Engineering, LGBTQIA+

1 Introduction

Personas are widely used to support design, requirements, and service decisions, as they help teams understand users and make goals, barriers, tasks, and usage contexts more concrete [5, 8]. Despite these benefits, building useful personas remains demanding, since manual construction may require time, domain knowledge, and careful interpretation of user-related evidence.

Large Language Models (LLMs) offer a promising alternative for making persona generation faster and more scalable [1, 2, 6]. However, faster generation does not automatically imply better personas. LLM-generated personas may be generic, weakly connected to the target domain, or affected by assumptions and stereotypes present in the model. These risks become more critical in diversity-related contexts. In this paper, inclusive personas refer to personas designed to support design and requirements activities involving marginalized or underrepresented groups. Their goal is not to represent an entire group, but to make socially situated needs, barriers, support networks, and interaction contexts visible while reducing the risk of generic, oversimplified, or stereotyped representations [3, 4, 7].

This tension between scalability, contextual adherence, and representational responsibility motivates the approach proposed in this paper. We investigate the generation of inclusive personas using LLMs supported by a structured document-based dataset and Retrieval-Augmented Generation (RAG). The main contribution is not merely applying RAG to persona generation. Rather, the document-based dataset is treated as a central grounding layer: documentary evidence about LGBTQIA+ experiences is collected, structured, categorized, and organized into retrieval units that provide contextual support for persona construction.

The approach compares two generation settings. In the document-grounded setting, the LLM receives evidence retrieved from the document-based dataset before generating the persona. In the no-dataset baseline setting, the same model generates personas from the same usage scenarios, but without access to retrieved evidence. This comparison allows us to investigate whether document-based grounding changes how personas are perceived as artifacts for reasoning about systems, services, and requirements, and as representations of people in diversity-related contexts.

Based on the proposed approach, this work seeks to answer the following research question:

RQ. How do inclusive personas generated by LLMs with document-based grounding (via RAG) compare to personas generated without such support in terms of perceived structural quality and sensitivity to diversity?

To answer this question, two personas were generated for each of three LGBTQIA+ usage scenarios: one in the document-grounded setting and another one in the no-dataset baseline setting. The personas were evaluated from two complementary perspectives: perceived technical/structural quality and sensitivity to diversity. This work provides three preliminary contributions: (i) a structured document-based dataset to support inclusive persona generation;

(ii) a RAG-based approach that compares document-grounded personas with a no-dataset baseline under the same scenarios and model configuration; and (iii) a preliminary evaluation instrument centered on technical/structural quality and sensitivity to diversity. The results suggest that document-based grounding can improve specificity and sensitivity to diversity in some scenarios, but does not automatically guarantee better personas when retrieved evidence is poorly aligned or unevenly represented in the dataset.

2 Related Work

Personas are established artifacts in design and requirements engineering. Salminen et al. [5] show, in a systematic review, that personas are used to represent user needs in diverse populations and to support design through empathy and communication, with software development appearing among the domains in which the technique is most frequently employed. In requirements engineering, Wang et al. [8] show that personas continue to be used in software projects, although their frequency, perceived effectiveness, and application challenges vary across companies, projects, and professional profiles. These works reinforce the relevance of personas as artifacts for reasoning about users, but they also show that their adoption and quality depend on how they are constructed and used.

When personas are used in diversity-related contexts, representation becomes a central concern. Marsden and Pröbster [3] argue that personas should not be built from reduced or isolated identities, since multiple identities need to be considered in interaction to avoid superficial and stereotyped representations. Similarly, Sun et al. [7] show that personas for people with complex needs may become “oversimplified or stereotypical” and emphasize the importance of data transparency, language, tone, and balance between abilities and constraints. These concerns are also consistent with broader evidence from software engineering: Olson et al. [4] show that marginalized communities’ ethical concerns about digital platforms often involve discrimination and misrepresentation. Together, these works suggest that inclusive personas require not only narrative coherence, but also contextual specificity and representational care.

Recent work has explored the use of LLMs to make persona creation faster and more scalable. Ayach et al. [1] investigate proto-persona generation through prompt engineering and report gains in efficiency, reduced effort, and improvements in perceived quality and reuse in later stages of the discovery process. However, they also point to limitations related to generalization and domain specificity. Other studies further highlight these risks. Salminen et al. [6] identify biases in LLM-generated personas related to age, occupation, pain points, and geographic context. Lazik et al. [2] observe that LLM-generated personas may be perceived as informative and consistent, while still tending to reproduce stereotypes and miss contextual or emotional insights.

Unlike these works, our study investigates the generation of inclusive personas with LLMs grounded in a document-based dataset and RAG. We compare personas generated with and without dataset support across LGBTQIA+ scenarios and evaluate them from two complementary perspectives: technical/structural quality and diversity. In doing so, we examine whether documentary grounding can

improve contextual specificity and sensitivity to diversity, while also identifying limitations introduced by retrieval quality and dataset coverage.

3 Proposed Approach

This section presents the proposed approach for generating and evaluating inclusive personas using a structured document-based dataset. The approach reduces exclusive dependence on the language model’s general knowledge by incorporating documentary evidence into persona generation.

The contribution is not merely the use of RAG as a generic retrieval mechanism. Instead, the document-based dataset is treated as a structured grounding layer: documentary evidence about the LGBTQIA+ experiences is transformed into categorized evidence records, enriched with metadata, organized into a RAG-ready retrieval index, and used to support persona-relevant dimensions such as needs, barriers, support networks, service interactions, and design opportunities.

The pipeline follows six steps: (i) selecting a textual corpus related to LGBTQIA+ experiences and diversity-related contexts; (ii) structuring it into evidence records; (iii) organizing these records into a RAG-ready retrieval index; (iv) generating personas in the document-grounded setting; (v) generating personas in the no-dataset baseline setting; and (vi) evaluating the produced personas. Figure 1 summarizes the pipeline.

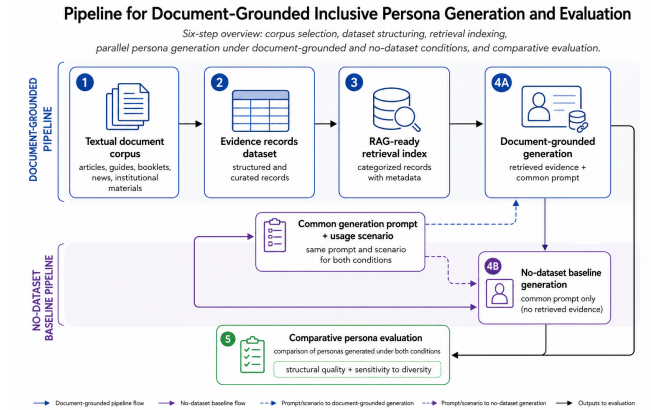


Figure 1: Overview of the proposed pipeline for document-grounded inclusive persona generation and evaluation.

In the document-grounded setting, the model receives the generation prompt with evidence retrieved from the dataset. In the no-dataset baseline setting, personas are generated with the same model and usage scenarios, but without retrieved evidence. This design enables comparison between both settings while partially controlling for the effects of the model and scenario.

3.1 Construction of the Document-Based Dataset

The corpus was compiled from 251 source documents, including academic articles, magazine and news materials, guides, manuals, booklets, government documents, and institutional materials

produced by Brazilian organizations, collectives, and public institutions connected to LGBTQIA+ agendas, such as IBRAT, ABGLT, and Aliança Nacional LGBTQIA+. The dataset focuses on LGBTQIA+ experiences in Brazil related to sexual and gender diversity, access to services, documentation, health, work, discrimination, family recognition, and support networks. Documents were included when they addressed these themes and excluded when they were outside the study scope, duplicated, noisy, inconsistent, or not useful for persona construction.

The dataset construction combined manual curation and LLM-supported extraction. A preliminary manual reading identified relevant excerpts and informed a seed layer of 128 manually curated records, which established standards for data structuring and information extraction. Then, the 251 source documents and the manually identified excerpts were processed through an LLM-supported extraction process, generating 4,649 candidate evidence records. After manual curation, 2,644 records were retained; the seed and silver layers were then consolidated and cleaned, resulting in 2,710 unique evidence records.

Thus, the corpus consisted of 251 source documents, while the final dataset contains 2,710 evidence records extracted, structured, and curated from those documents. After reclassifying records without a specific focal group as General LGBTQIA+, this category concentrated 1,602 records, followed by trans people, with 825 records. The remaining groups presented lower coverage: gay people, with 125 records; lesbians, with 112 records; and bisexual people, with 46 records. Figure 2 presents this distribution by primary focal group.

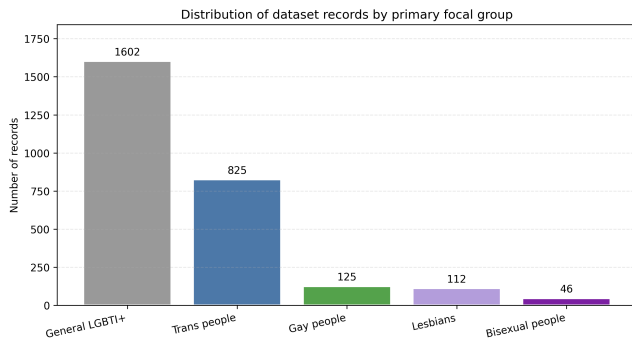


Figure 2: Distribution of dataset records by primary focal group.

This uneven coverage is relevant for interpreting the results: better-covered groups or themes may retrieve more specific and aligned evidence, while lower-coverage scenarios may retrieve more generic or weakly aligned records. For instance, the larger number of trans-related records may have supported the trans woman/government app scenario, whereas lower lesbian- and gay-specific coverage may have affected the lesbian mother and gay man/adoption scenarios.

Finally, each record was organized as a retrieval unit with controlled metadata, including primary focal group, domain, theme, subtheme, age group, trans-specific cut when applicable, target persona fields, source document, and page references. These records formed a RAG-ready retrieval index, allowing the system to select

evidence according to each usage scenario and metadata filters when applicable.

This structure differentiates the proposed approach from a generic RAG setup. The retrieved records are not arbitrary contextual passages; they are evidence units designed to inform persona-relevant dimensions such as needs, barriers, interactions with services, support networks, and design opportunities.

3.2 Persona Generation

Persona generation was conducted under two comparison settings: the document-grounded setting, in which the model was supported by the document-based dataset and RAG, and the no-dataset baseline setting, in which the same model generated personas without retrieved documentary evidence. All personas were generated with qwen2.5:7b-instruct and from the same usage scenarios.

In the document-grounded setting, the system retrieved relevant records from the RAG-ready index and inserted them into the anchor-generation prompt as documentary evidence. In the no-dataset baseline setting, personas were generated with the same model, scenario prompts, two-step generation strategy, and final persona-generation prompt, but without retrieved evidence. The prompt templates, scenario prompts, filters, model, and generation configuration are provided in the supplementary material.

By keeping the model, scenarios, and final persona-generation prompt constant across both settings, the comparison focuses on the effect of documentary support on contextual specificity, description of barriers, and sensitivity to diversity.

3.3 Scenarios Used

Persona generation was conducted from three usage scenarios intentionally selected to represent interactions with services, digital platforms, and institutional processes in which inclusion, diversity, access to rights, and family recognition could emerge. The scenarios were: (i) an adult trans woman using a government application for documentation and access to services; (ii) a lesbian woman, mother, and married to another woman, using applications and services to manage school, health, and family routines; and (iii) a gay man married to another man, in the process of adopting a child, dealing with documentation, public services, and family follow-up.

The goal was not to compare LGBTQIA+ identities with one another, but to observe whether different levels of documentary coverage could affect contextual richness, specificity of needs, description of barriers, support networks, interactions with services, and design opportunities. For each scenario, two personas were produced: one in the document-grounded setting and another in the no-dataset baseline setting, both starting from the same initial prompt.

4 Preliminary Study

The six personas evaluated in this study were generated using the pipeline described in Section 3. The study compares personas produced in the document-grounded setting and in the no-dataset baseline setting across the three usage scenarios previously described. Rather than repeating the generation pipeline, this section focuses on the evaluation design, participants, instrument, procedure, and analysis.

The exploratory evaluation aimed to identify how documentary grounding influenced evaluators’ perceptions of the personas as artifacts for design and requirements decisions and as representations of socially situated subjects. The evaluation crossed three usage scenarios with two generation settings, resulting in six final personas assessed from the perspectives of perceived structural quality and sensitivity to diversity.

4.1 Participants

Four evaluators participated in the study, selected through intentional sampling. The selection considered participants’ prior familiarity with the concept of personas and their ability to evaluate the artifacts from both a structural and a diversity-and-inclusion perspective.

The participants included two master’s students and two undergraduate students. All participants had some degree of familiarity with personas in contexts of design, requirements, or system evaluation. In addition, three participants self-identified as LGBTQIA+, which contributed perspectives especially attentive to representativeness, language, stereotypes, social plausibility, and sensitivity to the experiences portrayed. The participant profile brought together complementary backgrounds, allowing the personas to be examined both as artifacts supporting design and as representations of socially situated subjects and contexts.

Participants’ self-identification was reported voluntarily and was used only to characterize the evaluator group. The study did not compare responses according to participants’ identities, nor did it treat LGBTQIA+ participants as representatives of the communities portrayed in the personas. Their role in the study was to evaluate the generated artifacts from the perspective of persona quality, structural usefulness, and sensitivity to diversity.

4.2 Evaluation Instrument

The evaluation instrument was structured to operationalize the two dimensions stated in the research question: perceived structural quality and sensitivity to diversity. The perceived structural quality dimension evaluated the persona as a design and requirements artifact, considering internal coherence, specificity, plausibility, needs and barriers, and control of invention. The sensitivity to diversity dimension evaluated how the persona represented subjects, contexts, and experiences associated with LGBTQIA+ populations, considering cognitive empathy, affective empathy, behavioral empathy, and the absence of stereotypes.

To make the distinction between these two perspectives more explicit, Table 1 summarizes what each perspective evaluates, the criteria associated with it, and its relevance to the study.

The criteria were adapted from prior work on proto-personas generated with GenAI [1] and extended to the context of inclusive LGBTQIA+ personas. In addition to criteria-oriented evaluations, the instrument included fields for free-text comments, which allowed participants to record qualitative perceptions of the personas. The full instrument and criteria are provided in the supplementary material.

Table 1: Evaluation perspectives used in the preliminary study.

Perspective	Evaluation focus
Technical/structural quality	Evaluates the persona as an artifact for design and requirements reasoning. Criteria include internal coherence, specificity, plausibility, needs and barriers, and control of invention. This perspective indicates whether the persona can support design and requirements decisions.
Diversity sensitivity	Evaluates how the persona represents subjects, contexts, and experiences associated with LGBTQIA+ populations. Criteria include cognitive empathy, affective empathy, behavioral empathy, and absence of stereotypes. This perspective indicates whether the persona avoids flattened, stereotyped, or insensitive representations.

4.3 Procedure

The study was conducted through individual sessions. Participants received a brief explanation of the study goals, the evaluation task, and the use of the collected data for research purposes. Participation was voluntary, and participants evaluated generated artifacts rather than reporting personal life experiences. A supporting presentation introduced the concept of persona and the evaluation criteria, and an electronic form was used to present the personas, collect scores, and record free-text comments.

For each usage scenario, participants evaluated two final personas: one generated in the document-grounded setting and one in the no-dataset baseline setting. Personas were presented sequentially, one at a time, and participants were not informed about the generation setting. All participants followed the same fixed order: trans woman/government app; lesbian mother/family routine; and gay man/adoption process.

Personas followed a standardized textual format with descriptive and design-oriented fields, such as biography, needs, barriers, support networks, service interactions, inclusion sensitivities, and design opportunities. Participants read each persona, assigned ratings according to the evaluation criteria, and could add free-text comments.

The exact evaluation time was not logged; therefore, future replications should record time-on-task to help interpret evaluation effort. The collected ratings and comments were anonymized before analysis, and no identifying information is reported in the paper or supplementary material.

4.4 Data Analysis

The ratings were analyzed descriptively, considering the two evaluation perspectives and the comparison between generation settings. Given the exploratory nature of the study and the small number of evaluators, the analysis does not aim at statistical generalization. Instead, it identifies patterns in ratings and comments that indicate when documentary grounding may contribute to persona generation or introduce limitations.

Free-text comments complemented the ratings by helping interpret cases of contextual specificity, misalignment, excessive emphasis on sensitive themes, or representational concerns. Reported comments are anonymized using participant identifiers (P1–P4), and comments originally written in Portuguese were translated into English by the authors.

5 Results

5.1 Ordinal Rating

This section presents the results of the preliminary study based on 1–5 ordinal ratings and qualitative comments. The criteria covered perceived structural quality and sensitivity to diversity. Since the evaluation involved four participants, results should be interpreted as preliminary evidence rather than conclusive statistical evidence.

Figure 3 summarizes rating distributions by scenario and generation setting. Because the ratings are ordinal, we report the distribution, median (*Md*), and first and third quartiles (*Q1*–*Q3*). In the trans woman/government app scenario, the no-dataset baseline obtained *Md* = 4.0 (*Q1* = 3.0, *Q3* = 5.0), while the document-grounded setting obtained *Md* = 5.0 (*Q1* = 5.0, *Q3* = 5.0). In the lesbian mother/family routine scenario, the no-dataset baseline obtained *Md* = 5.0 (*Q1* = 4.0, *Q3* = 5.0), while the document-grounded setting obtained *Md* = 3.5 (*Q1* = 2.5, *Q3* = 4.0). In the gay man/adoption process scenario, both settings obtained *Md* = 5.0 (*Q1* = 4.0, *Q3* = 5.0).

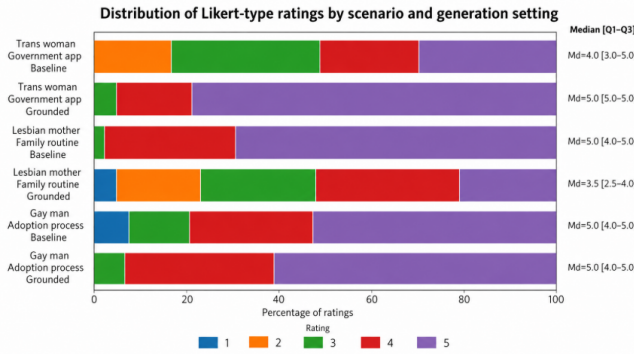


Figure 3: Ordinal rating distribution by scenario and setting.

Figure 3 shows that document-based grounding varied across scenarios: it was better evaluated in the trans woman/government app scenario, worse in the lesbian mother/family routine scenario, and equivalent in median in the gay man/adoption process scenario. In the lesbian mother/family routine scenario, lower ratings may be related to retrieval behavior: only 19 lesbian-specific records were associated with lesbian motherhood, while 80 were associated with violence and protection; five of the eight retrieved records for this scenario were associated with violence and protection. These data do not establish causality, but suggest a plausible relationship between retrieved evidence and the prominence of domestic violence in the generated persona.

The comments were originally written in Portuguese and translated by the authors. Participant identifiers (P1–P4) indicate different evaluators.

5.2 Perceptions by Scenario

In the trans woman/government app scenario, the document-grounded persona was perceived as more specific and useful for design decisions, especially regarding respectful service, social name, document updating, access to services, and fear of discrimination. The no-dataset baseline was considered plausible but more generic.

However, the document-grounded persona also raised a representational concern: one evaluator observed that “the text presents the deadname before the transition,” arguing that this practice may be offensive to trans people and irrelevant to the persona’s biography (P3). This suggests that greater specificity may also introduce risks of unnecessary exposure when sensitive information is not handled carefully.

In the lesbian mother/family routine scenario, the no-dataset baseline was perceived as more balanced, addressing motherhood, family routine, school, health, and technology use without reducing the character to her sexual orientation. In contrast, the document-grounded persona was considered less aligned with the scenario because of the recurring presence of domestic violence. One evaluator directly asked: “Does this person suffer domestic violence?” (P2). This illustrates how retrieved evidence related to violence and protection may have shifted the persona away from school, health, and family routine tasks.

In the gay man/adoption process scenario, both settings obtained the same median, but comments suggest different interpretations. The no-dataset baseline was coherent for discussing bureaucracy, documentation, and adoption follow-up, but some evaluators noted that it could represent almost any person in this process. One evaluator stated that, “without the diversity issue, [the persona] could represent anyone with adoption difficulties” (P3). By contrast, the document-grounded persona was perceived as more specific to gay fatherhood, incorporating prejudice, social stigma, support networks, legal guidance, and recognition of family diversity.

5.3 Discussion

The findings suggest that document-based grounding can support more specific and diversity-sensitive personas when the retrieved evidence is aligned with the design context. In the scenarios involving a trans woman using a government app and a gay man navigating an adoption process, grounding contributed to more situated descriptions of barriers, services, rights, documentation, and support needs. This indicates that retrieval can move an LLM-generated persona from a broadly plausible profile toward a more context-sensitive representation. However, the variation across scenarios shows that this benefit depends on the quality, coverage, and alignment of the evidence rather than being an automatic consequence of using RAG. This interpretation is consistent with prior work identifying both the usefulness of LLMs and their limitations regarding generalization and domain specificity [1].

The lesbian mother/family routine scenario illustrates the risks of poorly aligned retrieval. Because the retrieved records emphasized violence and protection, the generated persona shifted toward a sensitive issue that was not central to the original scenario. Documentary grounding is therefore not neutral: even credible evidence may produce a reductive persona when retrieval amplifies a narrow or thematically unbalanced set of records. The relevant question is not only whether a persona is grounded, but whether the evidence helps explain goals, barriers, service interactions, and design implications. This extends concerns about superficial, stereotyped, or oversimplified personas [3, 7].

The comparison with the no-dataset baseline further clarifies the contribution of grounding. In the adoption-process scenario,

the baseline persona was structurally coherent but less specific to LGBTQIA+ experiences, whereas the document-grounded persona made prejudice, family recognition, legal guidance, and support networks more visible. This contrast reinforces the need to evaluate inclusive personas through both structural and diversity-oriented lenses. A persona may be technically well formed while overlooking identity-specific barriers, or may include diversity-related themes in an excessive or misaligned way. This interpretation is consistent with prior findings that LLM-generated personas may appear informative while still missing contextual insights [2], as well as with concerns about diversity, discrimination, and misrepresentation in software systems [4, 7].

From a requirements engineering perspective, the approach should be treated as an early-stage support mechanism rather than a replacement for user research or human analysis. Generating grounded and ungrounded personas with the same model, prompt, and usage scenario allows reviewers to examine how retrieved evidence affects the representation of needs, barriers, and design opportunities. Human validation remains necessary before using these personas in design decisions, particularly when sensitive information is involved. Future tools should also expose the retrieved records, their provenance, and their influence on the generated persona. Future versions should further strengthen provenance metadata, document dates, versioning, periodic updates, coverage monitoring, and human review of sensitive, conflicting, outdated, or context-dependent evidence. Finally, because this exploratory study involved four evaluators and three usage scenarios, the results should be interpreted as initial indications rather than general performance claims.

6 Threats to Validity

Construct validity. The instrument was adapted from prior work and extended with diversity-oriented criteria, but may not capture all dimensions of inclusive persona quality. Since participants evaluated personas as artifacts, the study captures perceived quality and usefulness, not direct effectiveness in design or requirements tasks.

Internal validity. Scenario content, dataset coverage, retrieved evidence, retrieval behavior, and evaluator interpretation may have influenced persona perception. For example, in the lesbian family routine scenario, evidence related to violence and protection may have shifted the persona’s focus. Evaluation time was not recorded; future studies should log time-on-task to analyze effort and fatigue.

External validity. The study considered three LGBTQIA+ scenarios and four intentionally selected evaluators, which limits generalization. Findings should be interpreted as exploratory evidence, consistent with the IIER scope, rather than generalizable conclusions. Because the current dataset is centered on Brazilian LGBTQIA+ sources, it reflects a situated Brazilian perspective within the Global South rather than the full range of LGBTQIA+ experiences across geographic and sociocultural contexts. As future work, we plan to broaden the evidence base by incorporating documentary sources from additional countries and regions, while preserving the provenance and contextual specificity of each source. Future studies involving represented groups, sensitive identity-related

data, or real design use should undergo appropriate institutional ethics review, including CEP review in Brazil when applicable.

Conclusion validity. Results are descriptive and exploratory due to the small number of participants and the ordinal nature of the ratings. The qualitative comments contextualize the numerical patterns, but reflect a limited number of perspectives.

7 Conclusion

This paper presented a preliminary approach for generating inclusive personas with LLMs, a document-based dataset, and RAG. The approach uses documentary evidence about LGBTQIA+ experiences to support personas that are more contextually grounded than those relying only on the model’s general knowledge.

The preliminary study compared personas generated in the document-grounded setting and in the no-dataset baseline setting across three scenarios. Results suggest that document-based grounding may support greater specificity and sensitivity to diversity in some scenarios, but does not automatically guarantee superior quality. Its effect depends on the alignment between the usage scenario, dataset coverage, and retrieved evidence. Future work should expand participants and scenarios, include more diverse evaluator profiles, involve represented groups when appropriate, and refine retrieval and generation through thematic balancing, filtering of sensitive evidence, and better alignment between scenarios, records, and persona fields.

Artifact Availability

The supplementary material is available at: <https://doi.org/10.6084/m9.figshare.32284845>. It includes generated personas, prompt templates, generation scripts, sanitized RAG context metadata, the evaluation instrument, anonymized participant-level ratings, free-text comments, aggregated results, and figures. The full documentary corpus is not shared due to copyright restrictions and the sensitivity of LGBTQIA+ related materials; sanitized metadata and dataset documentation are provided for transparency and traceability.

Acknowledgments

We thank all the participants in the empirical study and USES Research Group members for their support. The authors also acknowledge the use of OpenAI’s ChatGPT, based on the GPT-5.5 Pro model, as a writing aid for linguistic refinement and figure generation in this manuscript. All generated content underwent critical review by the authors, who assume full responsibility for the originality, accuracy, and integrity of the final text, in accordance with CNPq’s Policy on Integrity in Scientific Activity (Ordinance No. 2.664/2026). This study was supported by CAPES - Financing Code 001, CNPq process 314797/2023-8, CNPq process 443934/2023-1, CNPq process 445029/2024-2, and Amazonas State Research Support Foundation - FAPEAM - through POSGRAD 2025-2026.

References

- [1] Fernando Ayach, Vitor Lameirão, Raul Leão, Jefferson Felizardo, Rafael Sobrinho, Vanessa Borges, Patrícia Matsubara, and Awdren L. Fontão. 2025. Generating Proto-Personas through Prompt Engineering: A Case Study on Efficiency, Effectiveness and Empathy. In *Anais do XXXIX Simpósio Brasileiro de Engenharia de Software (SBES 2025)*. Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil, 126–137. doi:10.5753/sbes.2025.9835

- [2] Christopher Klaus Lazik, Christopher Katins, Charlotte Kauter, Jonas Jakob, Caroline Jay, Lars Grunske, and Thomas Kosch. 2025. The Impostor is Among Us: Can Large Language Models Capture the Complexity of Human Personas?. In *Mensch und Computer 2025*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3743049.3743057
- [3] Nicola Marsden and Monika Pröbster. 2019. Personas and Identity: Looking at Multiple Identities to Inform the Construction of Personas. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300565
- [4] Lauren Olson, Emitza Guzmán Ortega, and Florian Kunneman. 2023. Along the Margins: Marginalised Communities’ Ethical Concerns about Social Platforms. In *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, Melbourne, Australia, 71–82. doi:10.1109/ICSE-SEIS58686.2023.00013
- [5] Joni Salminen, Kathleen Wenyun Guan, Soon Gyo Jung, and Bernard J. Jansen. 2022. Use Cases for Design Personas: A Systematic Review and New Frontiers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 543. doi:10.1145/3491102.3517589
- [6] Joni Salminen, Chang Liu, Wenjing Pian, Jianxing Chi, Essi Häyhänen, and Bernard J. Jansen. 2024. Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 510. doi:10.1145/3613904.3642036
- [7] Lipepei Sun, Tianzi Qin, Anran Hu, Jiale Zhang, Shuojia Lin, Jianyan Chen, Mona Ali, and Mirjana Prpa. 2025. Persona-L Has Entered the Chat: Leveraging LLMs and Ability-Based Framework for Personas of People with Complex Needs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 1109. doi:10.1145/3706598.3713445
- [8] Yi Wang, Chetan Arora, Xiao Liu, Thuong Hoang, Vasudha Malhotra, Ben Cheng, and John Grundy. 2025. Who Uses Personas in Requirements Engineering: The Practitioners’ Perspective. *Information and Software Technology* 178 (2025), 107609. doi:10.1016/j.infsof.2024.107609