Journal on Interactive Systems, 2023, 14:1 doi: 10.5753/jis.2023.3104

Investigating usability pitfalls in Brazilian and Foreign governmental chatbots

Mateus de Souza Monteiro [**Universidade Federal Fluminense**| ***mateusmonteiro@id.uff.br***]
Gabriely Oliveira da Silva Batista [**Universidade Federal Fluminense**| ***gabrielysilva@id.uff.br***]
Luciana Cardoso de Castro Salgado [**Universidade Federal Fluminense**| ***luciana@ic.uff.br***]

Abstract

The usability of chatbots is an essential aspect of e-government initiatives that aim to improve citizens' access to public services. Despite significant efforts to enhance e-gov usability in Brazil, there is still a lack of research examining usability pitfalls in human-chatbot interaction. Therefore, this study aims to fill this research gap by presenting an analysis and comparison of ten electronic government chatbots, five national (Brazilian) and five foreign (1 Argentine, 1 Portuguese, 1 American, 1 British, and 1 Singaporean), using an adapted version of Heuristics for chatbots. This study examines the design issues and opportunities that can affect the usability and adoption of e-government chatbots in Brazil. These include enhancing the navigation mechanism for dialogues, improving response times to requests, clearly indicating the end of the dialogue, providing guidance on inputting utterances correctly and implementing standardized utterances for the same actions.

Keywords: *Chatbot, Electronic Government, Usability, Service delivery*

1 Introduction

The concept of conversational agents has evolved and expanded since 1950 (Chaves and Gerosa, 2020). As defined by (Deibel and Evanhoe, 2021, pp. 31), chatbots refer to a solution that enables users to communicate through speech or text. These products can take various forms, such as chatbots, robots, multimodal agents, chatterbots, virtual humans, embodied conversational agents, and others. It is expected that the market for chatbots will grow by more than $100 billion USD by 2026 (Mordor, 2022).

Electronic government (e-gov) chatbots in Brazil serve primarily as an "ombudsman", centralizing and providing information on frequently asked questions. For instance, the Assis City Hall uses a chatbot named Sissa, which operates via WhatsApp and serves as a guide to provide public service information related to health, environment, service availability, and other relevant topics.

Despite significant growth in development and deployment, current chatbot applications face significant challenges (Brandtzaeg and Følstad, 2018). According to Wang et al. (2022), the primary goal of chatbots in e-gov is to enhance the quality of services provided to citizens, rather than simply introducing new technology. Besides, there have been widespread complaints and dissatisfaction regarding chatbots on government portals (Wang et al., 2022).

This poses a significant obstacle to adoption, especially given the already challenging nature of engagement in e-government solutions (Teo et al., 2008). Improving the quality of service delivery is crucial for building positive relationships with citizens (Monteiro et al., 2021), rather than solely focusing on growth and cost reductions, as is the case in companies (Huai, 2011).

One alternative approach to measuring the quality of e-gov chatbots is to evaluate and characterize their interaction quality, as demonstrated in Zaidi and Qteishat (2012). However, there is still a lack of research in the literature that comprehensively examines the usability problems of e-gov chatbots in Brazil. Understanding the digitalization of public services in Brazil is of special concern due to potential barriers to user adoption, such as social inequality and population aging (Pedrosa et al., 2022). This issue is particularly concerning given the prevalence of usability problems in solutions that are the first line of assistance for citizens seeking information and support.

In our previous study (da Silva Batista et al., 2022), we used Langevin's heuristic (Langevin et al., 2021) to identify usability problems in e-gov chatbots. Our findings revealed several usability issues with high severity ratings. While it is crucial to identify and assess the severity of usability issues (Langevin et al., 2021), the study (da Silva Batista et al., 2022) indicated the need for a more comprehensive investigation into the usability issues. In this extended version, we aim to explore potential design issues and opportunities that can improve e-gov chatbots and enhance the user experience in this critical domain of citizen services.

Considering the challenges above, this work has the following guiding questions:

- (RQ1) "What are the main usability problems in e-Gov chatbots?" and
- (RQ2) "What are the design issues for chatbot development?".

To answer the research questions, we inspected Brazilian and foreign e-gov chatbots. The study was divided into four steps:

- Preparation (S1): We selected ten chatbots, five Brazilian and five foreign (1 Argentine, 1 Portuguese, 1 American, 1 British, and 1 Singaporean), using five criteria. For each chatbot, we developed an adapted scenario that matched its objectives. For example, in the case of RioCard, a citizen might seek help with adjusting their card.

- Heuristic evaluation (S2): Using Langevin's heuristics (Langevin et al., 2021), we conducted a usability inspection of the user interface design to identify any usability problems. Langevin's approach maintains the same process as Nielsen's heuristics (Nielsen, 1992), but it focuses specifically on chatbots and includes additional heuristics and adaptations of existing ones.
- Comparison and Consolidation (S3): We compared the results of national and foreign chatbots.
- Conclusion (S4): Our study identifies important design issues that require improvement to enhance the usability of Brazilian e-gov chatbots, providing a starting point for discussion and technology enhancement in this domain.

The chatbots' inspection results reveal important design issues considering that those systems are tools to empower citizens to answer questions and requests: enhancing the mechanism to navigate through the dialogues; improving response time to requests; providing clear indications of the end of the dialogue; offering guidance on how to input utterances correctly; implementing standardized utterances. Those results provide a starting point for usability discussions in this domain with characteristics that might positively or negatively impact the engagement of human-chatbot interaction.

This article is divided into six sections, including the introduction. Section 2 discusses previous usability studies conducted on Brazilian e-government applications, while Section 3 focuses on works that discuss chatbots and usability. The methodology used in this study is presented in Section 4. In Section 5, the results of the study are presented. The discussions and limitations of the work are explored in Section 6. Finally, the article concludes with a summary of the main findings and their implications in Section 7.

2 Usability studies in Brazillian e-gov applications

Considering studies of e-gov usability in Brazil, for example, to overcome adoption in Brazillian Health Information Systems (HIS), Welchen et al. (2022) interviewed doctors to measure overall system evaluation and identify the impacts of such technologies in Brazil. To investigate, the authors used a National Usability-focused HIS Scale (NuHISS) that includes seven factors: technical quality, information quality, feedback, ease of use, internal collaboration, benefits, and inter-organizational collaboration. Besides validating the NuHISS in the Brazillian domain, results indicated only 38.9% were identified as high quality. Domingues et al. (2021) present a comparative map for telecare in Brazil. The authors summarize the key factors to promote interaction quality: ease of use and the number of clicks to perform an action.

Guimarães et al. (2014) explore the effects of e-governance transparency in Brazil. In Brazil, transparency is a civil right, allowing citizens to see if public funds are appropriately used, citizenship exercise, and strengthening relations between government and citizens. After applying Jakob Nielsen's heuristic evaluation (Nielsen and Molich, 1990) on 15 e-gov sites, results presented a direct link between governmental sites requiring transparency and needing more usability. As tools to promote citizenship, those systems should be useful and present interactivity, empowering citizens and promoting quality of life and social interaction. Similarly, de Moura Regly and Souza (2022) applied heuristic evaluation of the Transparency Portal, a governmental platform that allows citizens to search for information about the application of public resources.

Another challenge largely addressed for e-gov in Brazil is accessibility. In addition to making the usability of an application unfeasible, the accessibility of websites in Brazil is a law (Brasil, 2015). de Oliveira and Freire (2020) discuss usability problems related to the adoption of blind users. Santos et al. (2021) elaborate on the need for improving service delivery in e-government solutions for deaf users. Quispe and Eler (2018) feature 35 accessibility heuristics based on e-MAG, the Accessibility Model for E-Gov (from Portuguese: Modelo de Acessibilidade do Governo Eletrônico). According to the authors, using heuristics stimulates the dissemination of knowledge and the development of accessible platforms, the exercise of citizenship, and even the advancement of the services provided. Similarly, Carvalho et al. (2017) investigates the accessibility of e-government solutions. However, they present an evaluation process of several e-gov platforms.

While efforts have been made to enhance e-gov usability in Brazil, there is still a lack of research examining the usability of chatbots, which is the focus of this study. In the following section, we will present literature that delves into this topic.

3 Chatbots and usability

Artificial Intelligence (AI) has emerged as a promising solution to address various challenges faced by the public sector, such as inadequate resources and inefficiencies in e-government (Wang et al., 2022). For example, many local governments and departments in China are leveraging AI-powered chatbots to enhance service delivery and improve customer satisfaction. In their research, Wang et al. (2022) studied the initial adoption and post-adoption stages of chatbot implementation in China's local government using an analytical framework that considers both readiness and pressure factors. According to the authors, adopting chatbots is insufficient, chatbots should also perform well.

However, it is worth noting that most of the existing literature on chatbots and usability does not specifically focus on the domain of e-gov, but findings from various studies provide useful insights into factors affecting successful chatbot implementation, quality attributes that can serve as a checklist for implementation teams, and tools and methods for evaluating and improving chatbot usability, ultimately leading to better user experiences and outcomes. According to a recent systematic mapping on evaluation techniques for chatbot usability (Ren et al., 2019), the most commonly used methods are surveys, experiments, and usability tests.

For instance, in Jain et al. (2018), participants were asked to rate their experience using a Likert scale across metrics

such as ease of use, frustration, and provide feedback on the interface. In another study, the think-aloud technique was employed, with participants describing any usability problems encountered during interaction using their voices (Tielman et al., 2017). In Saenz et al. (2017), authors conducted a usability test of three chatbot platforms using the SUS (System Usability Scale) questionnaire combined with three other methods to rate participant feedback. However, despite those efforts to improve e-gov usability in Brazil, the literature still lacks to investigate the chatbots' usability, the focus of this research.

Borsci et al. (2022) provide toolkit for chatbot designers to use during the formative phase of chatbot development. The researchers first examined existing attributes from the literature, then conducted an online survey with chatbot designers and end-users to converge on a final corpus of attributes. Finally, they proposed the Bot Usability Scale (BUS), which is a five-dimensional scale designed to measure usability in chatbots. The dimensions of the BUS include perceived accessibility to chatbot functions, perceived quality of chatbot functions, perceived quality of conversation and information provided, perceived privacy and security, and time response.

Luk et al. (2022) explore the effectiveness of a chatbot in promoting COVID-19 vaccination in the context of vaccine hesitancy among young adults in Hong Kong. The researchers administered a post-intervention questionnaire that included the System Usability Scale (SUS). Their pilot study showed that using the designed chatbot resulted in a decrease in COVID-19 hesitancy among participants. In their study, Radziwill and Benton (2017) conducted a thorough literature review of academic and industry sources to identify 38 quality attributes that can be used to design effective conversational agents: efficiency, effectiveness, and satisfaction.

According to Langevin et al. (2021), creating a positive user experience with chatbots requires evaluation of both the conversation and the user interactions. To facilitate this process, the authors present a set of validated heuristics that researchers and practitioners can use during the formative evaluation of chatbots. These heuristics were adapted from Nielsen and Molich (1990) and take into account Grice's maxims of relevance and quality. The authors also introduce two new heuristics, Context Preservation and Trustworthiness. levels of user satisfaction.

To Langevin et al. (2021), providing a good user experience may require the evaluation of the conversation as well as use interactions. In their work Langevin et al. (2021), presented set of validated heuristics that researchers and practitioners could use during the formative evaluation of chatbots, adapting from heuristics (Nielsen and Molich, 1990). In line with Grice's maxims of relevance and quality, the authors introduce the heuristics Context preservation and Trustworthiness.

Although many studies concentrate on providing a list of attributes to enhance the formative evaluation and design of chatbots, none of them have analyzed and compared multiple agents in a particular domain, such as e-gov, which is the focus of our study. In the following section, we will present the methodology we used to accomplish this goal.

4 Methodology

The researchers responsible for the investigation were: R1, an entry-level undergrad HCI researcher with no background; R2, an HCI expert with 5+ years of experience in the field, with background in chatbots and culture; R3, an HCI expert and professor, with 10+ years of experience in the field, also with background in chatbots and culture.

- (S1) Step 1: Chatbot selection - collectively by R1, R2, and R3.
- (S2) Step 2: Heuristic Evaluation - conducted individually by R1 and R2, followed by the alignment of the results with R3.
- (S3) Step 3: Comparison and consolidation - conducted individually by R1 and R2, followed by the alignment of the results with R3.
- (S4) Step 4: Conclusion - Applied individually by R2, followed by the alignment of the results with R3.

We first (S1) conduced a chatbot selection. Similar to the process found in Höhn and Bongard-Blanchy (2021), we used the Google search engine through searches on sites that contained information about the active chatbots or academic work publications.

During the search, terms such as chatbot (and synonyms) were used and added with government, e-government, or e-gov. The search has been made in Portuguese, English, and Spanish. The chatbots should:

- i) Have the minimum of interaction without the need to insert confidential data,
- ii) Be active during the study
- iii) Be a government or public service provider.
- iv) Be active for users outside the country of origin, n case of foreign ones.

To investigate governmental chatbots, we applied Heuristic evaluation (S2), a set of usability heuristics commonly based on ten heuristics proposed by Nielsen in 1990 (Nielsen and Molich, 1990). Heuristics are established guidelines that, when incorporated into the design process, result in good interface design (Langevin et al., 2021). These heuristics help practitioners of Human-Computer Interaction (HCI) to identify usability problems that can affect users or slow down interaction, or cause unnecessary inconvenience to users (Nielsen and Molich, 1990).

Although widely used today, Nielsen and Molich's guidelines have yet to be adapted for the novel context of chatbots. For this reason, this paper uses Langevin's set of heuristics, a validated set of 11 heuristics adapted for conversational agents that can be generalized to text, voice, and multi-modal conversational agents (Langevin et al., 2021).

- (H1) Visibility of system status - The system should always keep users informed about what is going on, through appropriate feedback within a reasonable time, without overwhelming the user.
- (H2) Match between system and the real world - The system should use language, phrases, and concepts familiar to the user rather than system-oriented

terms or confusing terminology. Information should be presented in a natural and logical order, and include dialogue elements that create a smooth conversation through openings, mid-conversation guidance, and graceful exits.

- (H3) User control and freedom - Users should be able to effortlessly leave unwanted states without having to go through an extended dialogue. Support undo and redo functions.
- (H4) Consistency and standards - Users should not have to wonder whether different words, options, or actions mean the same thing. Follow platform conventions for the design of visual and interaction elements. Users should also be able to receive consistent responses even if they communicate the same function in multiple ways and modalities. Within the interaction, the system should have a consistent voice, style of language, and personality.
- (H5) Error prevention - Design conversations and interfaces to reduce the likelihood of problems occurring in the first place. Be prepared for pauses, conversation fillers, and interruptions, as well as dialogue failures, dead ends or sidetracks. Proactively prevent or eliminate potential error-prone conditions, and check and confirm with users before they commit an action.
- (H6) Help and guidance - The system should guide the user throughout the dialogue by clarifying system capabilities. Help features should be easy to retrieve and search, focused on the user's task, list concrete steps to be carried out, and not be too large. Make actions and options visible when appropriate.
- (H7) Flexibility and efficiency of use - Provide users with the appropriate (or preferred) input and output modality and hardware. Provide accelerators, such as command abbreviations, that are unseen by novices but speed up the interactions for experts, to ensure that the system is efficient.
- (H8) Aesthetic, minimalist and engaging design - Dialogues should not contain information which is irrelevant or rarely needed. Provide interactional elements that are necessary to engage the user and fit within the goal of the system. Interfaces should support short interactions and expand on the conversation if the user chooses.
- (H9) Help users recognize, diagnose and recover from errors - Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
- (H10) Context preservation - Maintain context preservation regarding the conversation topic intrasession, and if possible inter-session. Allow the user to reference past messages for further interactions to support implicit user expectations of conversations.
- (H11) Trustworthiness - The system should convey trustworthiness by ensuring privacy of user data, and by being transparent and truthful with the user. The system should not falsely claim to be human.

During the heuristic evaluation process, we collected and categorized all usability violations with evidence gathered from the interface to support our findings. Next, we assessed the severity of each violation to determine its impact on the user experience, considering the frequency, impact, and persistence of the problem. We refer to this as severity definition.

The severity of usability violations is divided into four degrees: first degree for cosmetic problems that do not affect the user's interaction with the system; second degree for minor problems with low repair priority; third degree for major problems that significantly affect critical usability factors, requiring high-priority repair; and fourth degree for catastrophic problems that, if not resolved, can prevent users from completing their tasks and achieving their goals. The evaluation process was applied on Brazilian and foreign chatbots, then compared (S3). Finally, in Conclusion (S4), every material collected and chatbots was revisited in order to find the major usability problems.

5 Results

5.1 S1 - Preparation

After a thorough selection process, we identified 25 chatbots as potential candidates for our study. However, we had to exclude some chatbots from our analysis mainly due to two criteria: (i) the need for confidential data and (ii) inactivity during the study period. Consequently, we focused our analysis on the remaining 10 chatbots that met our research requirements.

The national chatbots we examined included Mauá (C6), which belongs to the Federal Government; Rose (C7) and Sissa (C8) from the State Government; and SPTrans (C9) and Tomais (C10), both service providers for the Government. C8 provides assistance to users through WhatsApp for inquiries related to various topics, such as the environment, leisure pursuits, travel, education, and taxation. C7, on the other hand, assists citizens with scheduling appointments, public lighting, and taxes through its website.

Table 1. Final list of selected chatbots.

Id	Name	Channel	Country	Language
C1	DGAE	Website	Portugal	Portuguese
C2	Bot	WhatsApp and Website	Singapore	English
C3	Boti	WhatsApp	Argentine	Spanish
C4	Emma	Website	United States	English/Spanish
C5	TravelBot	Facebook	United Kingdom	English
C6	Maua	Website	Brazil	Portuguese-Br
C7	Rose	Website	Brazil	Portuguese-Br
C8	Sissa	WhatsApp	Brazil	Portuguese-Br
C9	Sp Transp	Facebook	Brazil	Portuguese-Br
C10	Tomais	WhatsApp	Brazil	Portuguese-Br

The Maua chatbot (C6) helps citizens register complaints, suggestions, compliments, and requests through its official website. C10 is the assistant for RioCardMais, a service provider for the state of Rio de Janeiro. Citizens can ask C10 questions about taxes, benefits, cancellations, and recharges among others via WhatsApp. C9 provides information about benefits for older adults, students, and services.

The foreign chatbots include four government-owned chatbots and one from a service provider, each from a different country. The United Kingdom offers C5, which helps with

metro lines, bus arrivals, and bus route status, among others. C5 uses Facebook Messenger as a platform for interaction. C1, from the General Economic Activities of Portugal, is a virtual assistant that answers questions related to economic activities, commerce, and services.

C4 provides citizenship and immigration services for the United States of America and can interact using English and Spanish. C3, through WhatsApp, provides services for the City Hall of Buenos Aires, the capital of Argentina. C2, from the Singaporean Government, through WhatsApp, acts as an FAQ assistant to help with questions related to Covid-19, including clarifications about fake news, vaccines, medical advice, health protocols, latest news, travel, and immigration, among others.

5.2 S2 - Heuristic evaluation

The first research question concerns the usability problems of e-gov chatbots. The following sections presents the results of the investigation (Table 2).

Table 2. Proportion of the severity of the usability violations found in each of the ten chatbots discussed in this articles, as well as the total number of heuristic violations in parentheses.

Chatbot (occurrence)	Langevin Heuristic (Langevin et al., 2021)									
	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
C1 (3)	*	*	2	*	*	*	*	1	*	3	*
C2 (1)	*	*	*	*	*	*	*	*	*	*	2
C3 (2)	3	*	*	*	*	*	*	*	3	*	*
C4 (2)	2	*	*	*	*	*	3	*	*	*	*
C5 (2)	*	*	4	*	4	*	*	*	*	*	*
C6 (3)	*	*	4	*	*	*	*	*	3	*	4
C7 (2)	1	*	*	*	*	2	*	*	*	*	*
C8 (5)	1	*	4	2	3	*	*	*	*	*	4
C9 (3)	*	*	4	3	*	*	*	*	*	4	*
C10 (4)	*	*	4	*	4	*	*	*	3	1	*

5.2.1 National chatbots

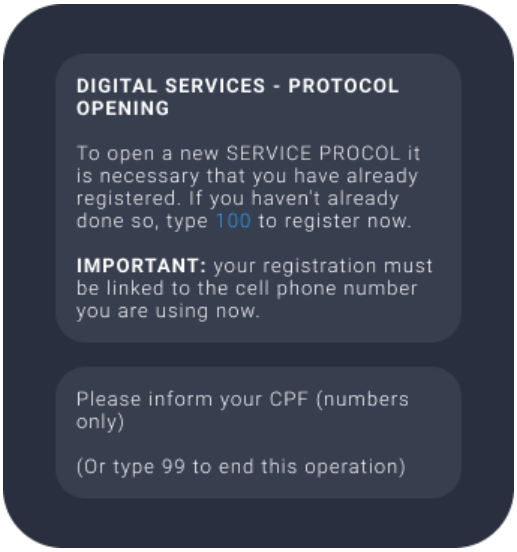

Figure 1. Sissa requests Personal Info.

National chatbots C6-C10 exhibit 17 violations of Langevin Heuristics, which raises concerns about the quality of e-gov service delivery in Brazil. The H3 heuristic, which pertains to user control and freedom, poses a significant usability problem for four out of five chatbots due to the challenge of dealing with unwanted states. This challenge is further complicated by the diverse educational and technological literacy levels of the Brazilian population. Heuristics 2, 7, and 8 do not present any violations.

C8 Sissa has the most severe usability violations among the chatbots. As it deals with citizen services, it requires personal information to complete requests but lacks transparency regarding user data privacy (H11). While C10 offers a link to user Terms of Use, Sissa lacks consistency (H4) and requires users to use different words to communicate the same function. Additionally, users cannot return to the previous menu since they can only move forward with the dialogues (Figure 1 depicts Sissa's collection of personal information).

5.2.2 Foreign chatbots

Foreign chatbots (C1-C5) have been found to violate nine of the Langevin Heuristics. Heuristics H1 and H3 were found to have recurring violations, while other heuristics (H5, H7-H11) only appeared once. However, all chatbots performed well on H2 (Consistency and Standards) and H4 (Error Prevention). There was no correlation found between the platform and the severity of the usability problems. It is noteworthy that even though C5 (TravelBot) did not have the most violations, it received the highest severity score of 4, indicating a significant usability problem that could impact the citizens of London. Given the importance of TravelBot's service, which includes bus arrival updates and service status, it is unreasonable to place the burden of error prevention on the user.

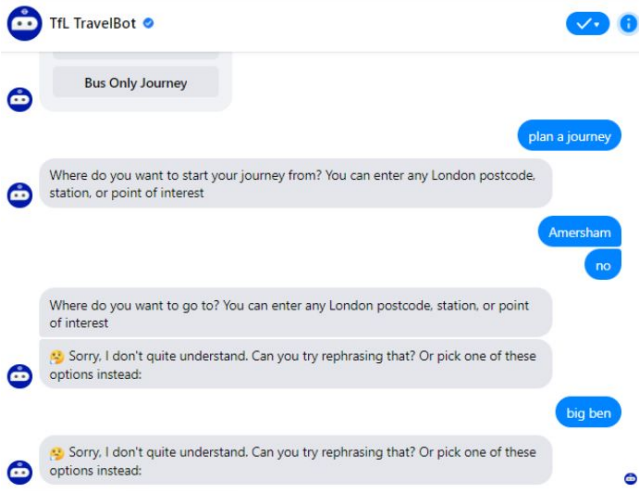

Figure 2. Travelbot trying to recover from an error.

An evaluator also identified an problem with the lack of support of the chatbot to navigate between dialog menus due to problems with error prevention and freedom (H5). In Figure 2, C5 attempts to recover from a typo. At a certain point in the conversation, C5 offers options, but the user decides to type, causing the dialog to enter a loop. Overall, foreign chatbots did not present many heuristics violations. However, the severity of the usability problems tended to present significantly severity ratings, which could significantly impact

the user experience.

5.3 S3 - Comparison and consolidation

In regards to the national chatbots, it is worth noting that only one bot did not violate H3. This heuristic is often associated with allowing users to navigate backwards and forwards within the dialogue without having to restart the entire conversation to return to a previous step. For instance, C5 provides persistent buttons to the user, but this feature is only available on Messenger as a platform for the bot.

C1 also presents a persistent button in the dialogue, but it does not necessarily facilitate navigation through the conversation. H3 also becomes relevant when recovering from an error (H5). For example, when the user tries to exit the current dialogue, C10 is unable to recover, requiring the user to restart the entire conversation. As Langevin suggests, designers should prepare the bot for pauses, fillers, and interruptions. For instance, if the bot is being used for urgent purposes, such as e-gov service delivery, major usability problems can cause users to waste time dealing with them.

Another important problem is trustworthiness (H11). As mentioned earlier, chatbots are systems that, like other solutions, can handle personal information. In the case of foreign chatbots, designers typically use their platform as the interface for the chatbot. In the case of national chatbots, the user's rights are only safeguarded through conversation, usually via WhatsApp. C10 provides a suitable alternative for this problem, as shown in Figure 3.

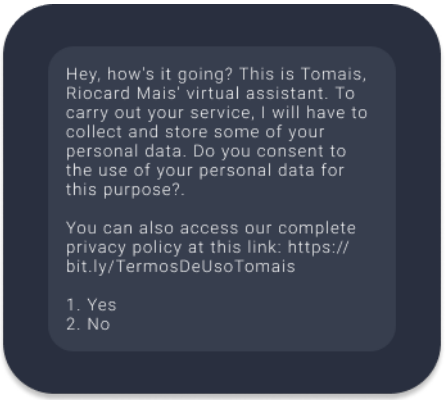

Figure 3. Tomais presents a link for the Terms of Use.

Conversational government agents frequently request sensitive information such as CPF (a unique identifier for Brazilian citizens used by the Internal Revenue Service), date of birth, telephone numbers, and email addresses. Therefore, it is crucial to present a clear security policy and inform users how their data will be used to encourage their confidence in the application. In the case of foreign chatbots, C2 harms users' trust by not being transparent about being an automated system. Although C2 does not falsely claim to be human, it is good practice to avoid omitting this information, as in the case of C10.

5.4 S4 - Conclusion

The second research question focuses on identifying design opportunities for e-gov chatbots based on the analysis of usability pitfalls. Through the analysis, five design issues (I) were identified among the chatbots to avoid pitfalls:

- I1 - Enhancing the mechanism to navigate through the dialogues.
- I2 - Improving response time to requests.
- I3 - Providing clear indications of the end of the dialogue.
- I4 - Offering guidance on how to input utterances correctly.
- I5 - Implementing standardized utterances.

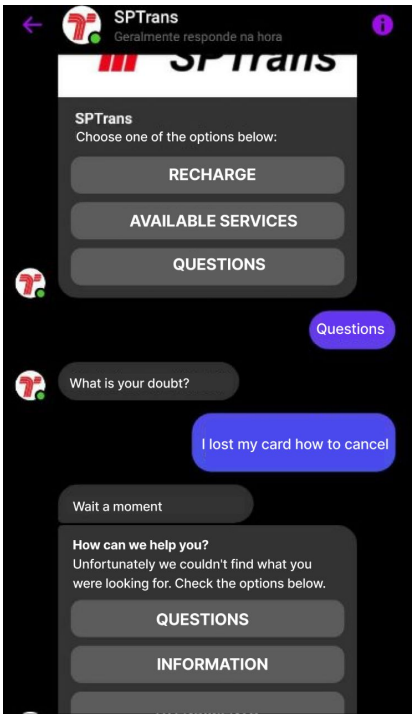

Figure 4. SPTrans (C9) informs the user that the action will take longer than usual.

Some chatbots offer persistent menus or buttons to allow users to go back or forward to different parts of the conversation, but on platforms like WhatsApp, this feature is not yet implemented. Having an efficient navigation mechanism is essential for users to quickly find the information they need and complete tasks within the chatbot. By implementing a persistent dialogue strategy or providing persistent menus or buttons, designers can improve the usability and overall user experience of e-government chatbots, which can ultimately lead to increased adoption and satisfaction among users (I1 - Enhancing the mechanism to navigate through the dialogues).

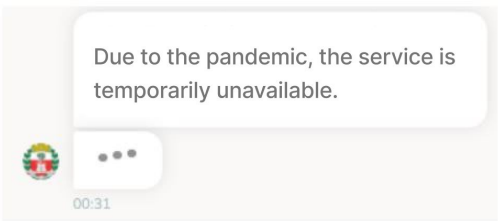

Figure 5. Rose (C7) says to user that the chatbot is still "Typing".

Delays in answering might occur due to errors or because the task requires more time to process. In such cases, error messages informing users that the task might take longer than usual (or compared to other tasks) might help users understand the system's current state (H1). Reducing delays in answering user requests (I2 - Improving response time to requests) is crucial for maintaining user engagement and satisfaction with the chatbot. However, in cases where this is not possible, using strategies such as error messages or icons that represent loading or typing, designers can help users understand the chatbot's current state and reduce frustration caused by delays in answering requests, ultimately leading to improved user satisfaction and adoption of e-government chatbots.

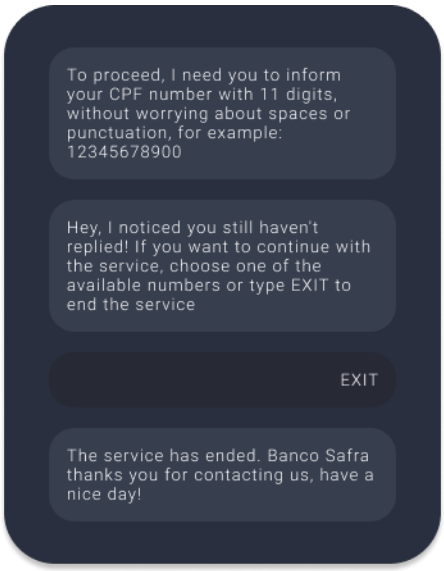

Figure 6. Chatbot Safra [1] informs the user how to properly enter an information.

Another pitfall that can impact the usability of e-government chatbots is the lack of clear feedback on the dialogue state, specifically at the end of the conversation. Some chatbots do not provide appropriate signals to inform the user that the dialogue has ended, leading to confusion and frustration. To address this problem, designers can follow (I3 - Providing clear indications of the end of the dialogue), using clear and concise messages such as "Thank you for your request, we will get back to you shortly" or "Is there anything else I can help you with?" to signal the end of the conversation. For example, in Figure 6, the Safra Bank chatbot uses the message "The service has ended. Banco Safra thanks you for contacting us, have a nice day!" to inform the user that the conversation has concluded.

Lack of guidance on how to properly interact with a chatbot can negatively impact usability. Without clear instructions, users may struggle to navigate the chatbot and input information or perform actions correctly. This can lead to errors, frustration, and ultimately, a negative user experience. To address this issue (I4 - Offering guidance on how input utterances correctly), designers can provide clear and concise guidance on how to operate common tasks within the chatbot, including using descriptive labels, placeholders, and examples (Figure 7). By doing so, users can better understand

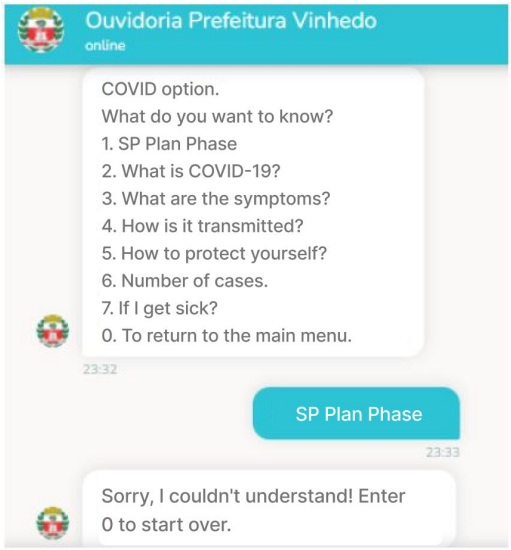

Figure 7. Rose lacks to present the correct way of entering the information.

the specific format or syntax required to input information or take actions within the chatbot, reducing errors and improving the overall user experience.

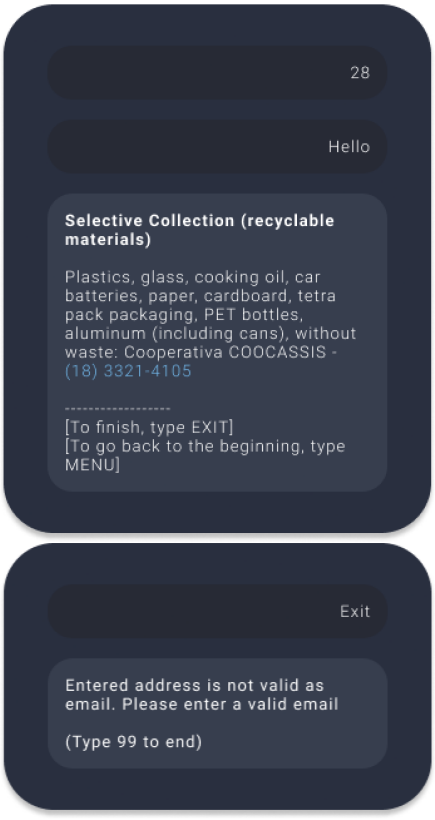

Figure 8. The chatbot requests different words for the same function. On the top, 'Exit' is used to end the chat, while on the bottom, '99' serves the same function.

Finally, a pitfall takes place when multiple actions with the same meaning are presented, causing redundancy and confusion for users. This violates the heuristic of consistency (H2) and can lead to user errors. To avoid this issue, designers should aim to use consistent terminology and actions throughout the chatbot (I5 - implementing standardized ut-

terances). In cases where multiple actions are necessary, designers should clearly explain the differences between each action. Figure 8 illustrates an example of this issue, where the same action is presented with different labels without clear explanation to the user.

6 Discussions and limitations

The chatbots' ability to address user needs effectively and efficiently is a well-established concept in the literature (Brandtzaeg and Følstad, 2018). However, this article draws attention to the violation of heuristic H3, which prevents users from navigating back and forth within a conversation, in the chatbots that were investigated. Furthermore, the findings underscore the significance of trustworthiness (H11) and urge designers to clearly communicate security policies and inform users about data usage. According to Borsci et al. (2022), trust in chatbots is primarily based on their ability to provide information and help users achieve their goals (i.e., credibility) rather than an unspecified sense of trustworthiness. In complex contexts, Luk et al. (2022) proposes a model comprising Confidence, Complacency, and Convenience to promote trust during the design process. The authors argue that the context already implies a lack of trust in the information provided by the bot, such as vaccines. In the case of the investigated chatbots, trust is linked to users' perception of privacy and transparency, which represents a specific type of trust rather than the credibility of the information provided.

According to Wang et al. (2022), the performance of chatbots becomes increasingly important for citizens after the initial adoption phase, and the level of performance can be correlated with the financial investment made in the chatbot development. Although our results are based on a limited number of chatbots, we can observe that the difference between national chatbots and those from developed countries is not significant in some cases, such as in C5. The choice of communication channel for chatbots used in e-gov services varies depending on the country. While foreign chatbots tend to prefer websites for their visual-centric conversations, Brazilian chatbots predominantly use WhatsApp due to its popularity among users (Höhn and Bongard-Blanchy, 2021). This preference for WhatsApp may also be due to the fact that it allows for secure and private conversations between citizens and the government.

When designing chatbots for e-government services, it is crucial to prioritize citizen needs and preferences to establish positive user experiences and build trust between citizens and the government (Borsci et al., 2022). As argued by Radziwill and Benton (2017), quality attributes are generally related to effectiveness (the accuracy and completeness with which specific users achieve their goals), efficiency (the extent to which resources are applied to achieve those goals), and satisfaction. It is not surprising that the investigated Brazilian chatbots exhibit a strong lack of efficiency with the need for enhancing the mechanism to navigate through the dialogues (I1) and the need for improving response time to requests (I2). Additionally, contextual variables in Brazil (Pedrosa et al., 2022) highlight the challenges associated with I3, I4, and I5, which directly affect lay-users and may reduce user satisfaction, thereby contributing to the overall dissatisfaction with government portals and negatively impacting the adoption of current and future technologies.

One of the limitations of our study is that we relied solely on inspection methods. While these methods offered valuable insights into the system's usability pitfalls, the absence of direct user involvement restricted our ability to capture their subjective experiences and perspectives. Currently, we are exploring alternative approaches to involve users in the re-design process. As a preliminary step, we engaged 28 HCI junior students in a discussion using the MoLIC (Modeling Language for Interaction as Conversation) (Prates and Barbosa, 2007) to address concerns in C8. These participants had no prior knowledge of the five identified pitfalls. However, most participants independently identified issues I1, I3, I4, and I5. For example, when presented with a potential solution for I1, participants made observations such as "Now the user can directly return to the desired menu, saving time (P2)" or "[...] the user no longer gets lost after selecting a menu option, enabling them to perform additional actions (P6)". While we have already begun investigating C8, our next step involves expanding our research to understand better how these pitfalls impact user adoption, a significant concern in e-gov solutions (Wang et al., 2022).

7 Conclusion

We presented a heuristic evaluation aimed at answering (RQ1) "What are the main usability problems in e-Gov chatbots?" and (RQ2) "What are the design issues for chatbot development?". Regarding RQ1, using Langevin heuristics (Langevin et al., 2021), it was possible to identify the main usability problems of the ten selected e-gov chatbots. Chatbots have the potential to improve e-government services by providing seamless and efficient user experiences. However, the success of chatbots in achieving this goal depends on various factors such as design, user needs, and trustworthiness. The investigated Brazilian chatbots demonstrate a lack of efficiency, heuristics violations, issues with user satisfaction and trust, highlighting the need for improvements in chatbot design for e-government services. Ultimately, well-designed and effective chatbots have the potential to improve e-government services and improve the relationship between citizens and their governments. Through our investigation of RQ2, we have identified five key design issues that have the potential to impact the usability and adoption of chatbots in e-government services. While these design issues serve as a helpful starting point for chatbot practitioners, further research is needed to fully understand the impact of user engagement and adoption.

Notes

This manuscript is an extended version of the work "Investigando chatbots governamentais: um panorama sobre a usabilidade dentro e fora do Brasil", da Silva Batista et al. (2022), presented in the Anais do X Workshop de Computação Aplicada em Governo Eletrônico - WCGE 2022.

References

Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D., and Chamberlain, A. (2022). The chatbot usability scale: the design and pilot of a usability scale for interaction with ai-based conversational agents. *Personal and Ubiquitous Computing*, 26:95–119.

Brandtzaeg, P. B. and Følstad, A. (2018). Chatbots: changing user needs and motivations. *interactions*, 25(5):38–43.

Brasil (2015). Lei nº 13.146, de 6 de julho de 2015. institui a lei brasileira de inclusão da pessoa com deficiência (estatuto da pessoa com deficiência). *Diário Oficial [da] República Federativa do Brasil*.

Carvalho, V., Cagnin, M., and Paiva, D. (2017). Avaliação de acessibilidade de web sites de governos estaduais do brasil. In *Anais do XIII Simpósio Brasileiro de Sistemas de Informação*, pages 116–123. SBC.

Chaves, A. P. and Gerosa, M. A. (2020). How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, pages 1–30.

da Silva Batista, G. O., de Souza Monteiro, M., and de Castro Salgado, L. C. (2022). Investigando chatbots governamentais: um panorama sobre a usabilidade dentro e fora do brasil. In *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*, pages 73–84. SBC.

de Moura Regly, T. R. and Souza, R. F. (2022). Disponibilização de dados e democratização do acesso à informação pública: uma análise do portal da transparência do governo federal. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 27.

de Oliveira, F. E. and Freire, A. (2020). Fatores que influenciam a aceitação e intenção de uso de governo eletrônico por pessoas com deficiência visual: um estudo sobre o site da receita federal. In *Anais Estendidos do XVI Simpósio Brasileiro de Sistemas de Informação*, pages 125–125. SBC.

Deibel, D. and Evanhoe, R. (2021). *Conversations with Things: UX design for Chat and Voice*. Rosenfeld Media.

Domingues, M. C. M., Melo, K. M., Miranda, A. F., de Freitas Cayres, A. Z., and Elias, R. (2021). Programa autoestima: Uma ferramenta web pública e amigável que integra formação de profissionais da saúde e acolhimento psicossocial da população. In *Anais do IX Workshop de Computação Aplicada em Governo Eletrônico*, pages 251–258. SBC.

Guimarães, C., Silva, L. D. R., Antunes, D. R., and Formighieri, R. (2014). E-governance transparency in brazil–the lack of usability is detrimental to citizenship. In *International Conference on Human Interface and the Management of Information*, pages 590–601. Springer.

Höhn, S. and Bongard-Blanchy, K. (2021). Heuristic evaluation of covid-19 chatbots. In *International Workshop on Chatbot Research and Design*, pages 131–144. Springer.

Huai, J. (2011). Quality evaluation of e-government public service. In *2011 International Conference on Management and Service Science*, pages 1–4. IEEE.

Jain, M., Kota, R., Kumar, P., and Patel, S. N. (2018). Convey: Exploring the use of a context view for chatbots. In *Proceedings of the 2018 chi conference on human factors in computing systems*, pages 1–6.

Langevin, R., Lordon, R. J., Avrahami, T., Cowan, B. R., Hirsch, T., and Hsieh, G. (2021). Heuristic evaluation of conversational agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Luk, T. T., Lui, J. H. T., and Wang, M. P. (2022). Efficacy, usability, and acceptability of a chatbot for promoting covid-19 vaccination in unvaccinated or booster-hesitant young adults: Pre-post pilot study. *Journal of Medical Internet Research*, 24(10):e39063.

Monteiro, M., Vasconcelos, L., Viterbo, J., Salgado, L., and Bernardini, F. (2021). Assessing the quality of local e-government service through citizen-sourcing applications. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 1178–1183. IEEE.

Mordor (2022). *CHATBOT MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2022 - 2027)*. Mordor Intelligence. https://www.mordorintelligence.com/pt/industry-reports/chatbot-market.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380.

Nielsen, J. and Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256.

Pedrosa, G. V., Judice, A., Judice, M., Araújo, L., Fleury, F., and Figueiredo, R. (2022). Applying user-centered design on digital transformation of public services: A case study in brazil. In *DG. O 2022: The 23rd Annual International Conference on Digital Government Research*, pages 372–379.

Prates, R. O. and Barbosa, S. D. J. (2007). Introdução à teoria e prática da interação humano computador fundamentada na engenharia semiótica. *Atualizações em informática*, pages 263–326.

Quispe, F. E. M. and Eler, M. M. (2018). Recomendações de acessibilidade para aplicativos móveis: uma contribuição para os padrões do governo digital brasileiro. In *Anais do XIV Simpósio Brasileiro de Sistemas de Informação*, pages 528–535. SBC.

Radziwill, N. M. and Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579*.

Ren, R., Castro, J. W., Acuña, S. T., and de Lara, J. (2019). Evaluation techniques for chatbot usability: A systematic mapping study. *International Journal of Software Engineering and Knowledge Engineering*, 29(11n12):1673–1702.

Saenz, J., Burgess, W., Gustitis, E., Mena, A., and Sasangohar, F. (2017). The usability analysis of chatbot technologies for internal personnel communications. In *IIE Annual Conference. Proceedings*, pages 1357–1362. Institute of Industrial and Systems Engineers (IISE).

Santos, S. K. S., Mendes, N. F. O., and da Silva, Q. P.

(2021). Propostas para o atendimento às pessoas surdas em serviços de governo eletrônico. In *Anais do IX Workshop de Computação Aplicada em Governo Eletrônico*, pages 1–12. SBC.

Teo, T. S., Srivastava, S. C., and Jiang, L. (2008). Trust and electronic government success: An empirical study. *Journal of management information systems*, 25(3):99–132.

Tielman, M. L., Neerincx, M. A., Bidarra, R., Kybartas, B., and Brinkman, W.-P. (2017). A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. *Journal of medical systems*, 41(8):1–10.

Wang, Y., Zhang, N., and Zhao, X. (2022). Understanding the determinants in the different government ai adoption stages: Evidence of local government chatbots in china. *Social Science Computer Review*, 40(2):534–554.

Welchen, V., Matte, J., Giacomello, C. P., Dalle Molle, F., and Camargo, M. E. (2022). Usability perception of the health information systems in brazil: the view of hospital health professionals on the electronic health record. *RAUSP Management Journal*, 57:264–279.

Zaidi, S. F. H. and Qteishat, M. K. (2012). Assessing e-government service delivery (government to citizen). *International journal of ebusiness and egovernment studies*, 4(1):45–54.