

# Investigando chatbots governamentais: um panorama sobre a usabilidade dentro e fora do Brasil

Gabriely Oliveira da Silva Batista<sup>1</sup>, Mateus de Souza Monteiro<sup>1</sup>, Luciana Cardoso de Castro Salgado<sup>1</sup>

<sup>1</sup>Instituto de Computação – Universidade Federal Fluminense (UFF)  
Niterói – RJ – Brazil

{gabrielysilva,mateusmonteiro}@id.uff.br, luciana@ic.uff.br

**Abstract.** *Electronic government applications usually associate their chatbots with problem-solving and answering citizen questions. This work aimed to investigate government chatbots from both inside and outside Brazil, looking for usability issues using a set of Nielsen-adapted heuristics designed for conversational devices. Ten chatbots were evaluated, five national (Brazilian) and five foreign (1 Argentine, 1 Portuguese, 1 American, 1 British, and 1 Singaporean). Results indicate a need for improvement in several heuristics issues in the service delivered on those platforms.*

**Resumo.** *Os aplicativos de governo eletrônico geralmente associam seus chatbots à solução de problemas e à ouvidoria de perguntas dos cidadãos. Este trabalho teve como objetivo avaliar a usabilidade de chatbots governamentais nacionais e internacionais, por meio de um conjunto de heurísticas voltadas para agentes conversacionais adaptadas das heurísticas de Nielsen. Foram avaliados dez chatbots, sendo cinco nacionais (brasileiros) e cinco estrangeiros (1 argentino, 1 português, 1 americano, 1 britânico e 1 singapurense). Os resultados indicam a necessidade de melhorias em diversas questões heurísticas no serviço prestado nessas plataformas.*

## 1. Introdução

Agentes inteligentes, assistentes digitais, agentes conversacionais ou simplesmente *chatbots*, objetivam promover a conversa entre humano e máquina em linguagem natural escrita ou falada [Dale, 2016]. No Brasil, os exemplos do uso de *chatbots* no domínio de e-gov (governo eletrônico) são diversos, sendo, na maioria dos casos, voltados para o papel de ouvidoria ou como um canal para centralizar as informações. Na Prefeitura de Assis, por exemplo, os cidadãos contam com o atendimento de uma assistente virtual, a Sissa<sup>1</sup>, a qual atua como uma ouvidoria para temas variados como, por exemplo, lazer, turismo, saúde e impostos; de maneira similar, a Rose<sup>2</sup> atua para a Prefeitura de Vinhedo. Já o *chatbot* Mauá<sup>3</sup>, também faz o papel de ouvidoria, mas para o Ministério da Infraestrutura.

Entretanto é necessário pontuar alguns desafios de atuação dessas tecnologias. A literatura aponta a complexidade dos *chatbots* em atenderem às expectativas dos usuários finais [Brandtzaeg e Følstad, 2018]. As tecnologias voltadas para o e-gov, em particular, apresentam desafios de engajamento e retenção de usuários [Teo, Srivastava,

---

<sup>1</sup> <https://www.assis.sp.gov.br/pagina/47/atendimento/zap-da-prefeitura>

<sup>2</sup> <https://www.vinhedo.sp.gov.br/>

<sup>3</sup> <https://www.gov.br/infraestrutura/pt-br/ouvidoria/conheca-o-mau>

e Jiang, 2008]. Além disso, em uma pesquisa recente da Take Blip<sup>4</sup>, empresa de grande porte desenvolvedora de chatbots, indicou que as principais reclamações apontadas pelos clientes estão associadas à usabilidade [Ventura, 2021]. Tais desafios podem impactar a qualidade do serviço oferecido por meio do *chatbot*.

De maneira geral, a boa qualidade na entrega de serviços resulta na redução de custos, crescimento e satisfação de clientes [Huai, 2011]. Quando o assunto é serviço público, os objetivos estão menos voltados aos ganhos financeiros [Monteiro et al., 2021]. Uma maneira de medir a qualidade da entrega do serviço oferecido aos cidadãos é investigar a qualidade da interação das pessoas com a tecnologia utilizada como meio, como visto em Zaidi e Qteishat (2012).

Neste sentido, este trabalho tem como questão de pesquisa norteadora “Quais são os potenciais impactos dos padrões de usabilidade violados em chatbots governamentais brasileiros e estrangeiros?”. Como objetivo, o de investigar a qualidade das soluções de *chatbots* brasileiros no domínio governamental com foco na usabilidade, perspectiva pouco explorada nos trabalhos encontrados na literatura. Além disso, comparar os resultados com *chatbots* de mesmo domínio que foram implantados fora do Brasil. Para tanto, foi utilizado um método de inspeção chamado de Avaliação Heurística [Nielsen, 1995; Nielsen e Molich, 1990], no qual os avaliadores analisam uma interface seguindo um conjunto de princípios com o objetivo de encontrar falhas de usabilidade. No entanto, foi utilizado um conjunto de heurísticas voltadas para agentes conversacionais adaptadas das heurísticas de Nielsen [Langevin, 2021].

Os principais resultados indicam um alto índice de problemas catastróficos nos *chatbots* brasileiros. Já os chatbots estrangeiros apresentaram um baixo índice de problemas nesse nível, e valores consideráveis em problemas de menor impacto. Este é um dado especialmente preocupante considerando que os chatbots atuam como fronteiras no atendimento entre governo e cidadão, em especial no caso dos chatbots encontrados no Brasil.

O trabalho está organizado em 4 seções além da introdução. Na Seção 2 é apresentado os Trabalhos relacionados. Na Seção 3 é exposto a Metodologia utilizada para o estudo. Depois, na Seção 4, os Resultados. Por fim, na Seção 5, as Conclusões.

## **2. Trabalhos relacionados**

Os trabalhos publicados nos últimos 5 anos sobre qualidade da interação em e-gov analisam e propõem melhorias de usabilidade e acessibilidade em websites ou aplicativos, mas não exploram o uso de chatbots. Por exemplo, Domingues et al. (2021) apresentam um mapa comparativo sobre as ferramentas utilizadas para Teleatendimento, considerando o contexto do acolhimento psicossocial da população. Dentre os fatores considerados, os autores analisam a facilidade de uso da plataforma, quantidade de clicks, dificuldade na criação de link para os atendimentos, e por fim a dificuldade de acesso da população em termos de registro, instalação da plataforma e suporte a integração.

A acessibilidade é um tema recorrente de investigação da qualidade da interface em plataformas governamentais. Oliveira e Freire (2020) analisam fatores como a aceitação e intenção de uso dos cidadãos com deficiência visual em plataformas de

---

<sup>4</sup> <https://www.take.net/>

e-gov. Dentre os fatores, os autores analisam com base na confiança, ansiedade e habilidade com tecnologia.

Já no trabalho de Santos, Mendes e Silva (2021), é apresentada uma discussão sobre a importância do atendimento às pessoas surdas brasileiras em plataformas de e-gov. Além disso, apresentam propostas para adequar e melhorar o atendimento e oferta desses serviços. Em Siqueira e Freire (2020) são investigados fatores comportamentais de cidadãos com deficiência na apresentação de denúncias sobre barreiras de acessibilidade em portais governamentais.

Já Carvalho, Cagnin e Paiva (2017) apresentaram a avaliação de acessibilidade de plataformas de e-gov estaduais. Por fim, Quispe e Eler (2018) apresentam um conjunto de 35 recomendações de acessibilidade para aplicativos móveis em e-gov adaptados da e-MAG (Modelo de Acessibilidade do Governo Eletrônico). De acordo com os autores, o desenvolvimento de plataformas acessíveis estimula a disseminação do conhecimento, exercício da cidadania e ainda o avanço dos serviços prestados. Em um outro tipo de análise acerca da qualidade de interface oferecidas pelas plataformas de e-gov, Pereira et al. (2021) analisam a hierarquização das informações e apresentam uma classificação taxonômica das terminologias dos serviços públicos ofertados em aplicações digitais.

De maneira geral, os trabalhos encontrados na literatura fornecem um cenário para acessibilidade em websites e aplicativos em e-gov. Para esta pesquisa, os trabalhos encontrados trazem um importante conjunto de informações sobre a aplicação de métodos de inspeção em e-gov. Como no caso de Carvalho, Cagnin e Paiva (2017), no qual, embora os portais estivessem no mesmo domínio, cada página foi avaliada considerando as peculiaridades de cada portal. Informação significativa a ser considerada quando avaliado serviços de diferentes localidades e prestadores de serviços.

Por fim, existe uma lacuna nos trabalhos a respeito da análise de *chatbots* dentro do contexto de e-gov. Tendo este trabalho não somente o objetivo de explorar este tema, mas também de apresentar um panorama atual destes recursos oferecidos em plataformas governamentais.

### **3. Metodologia**

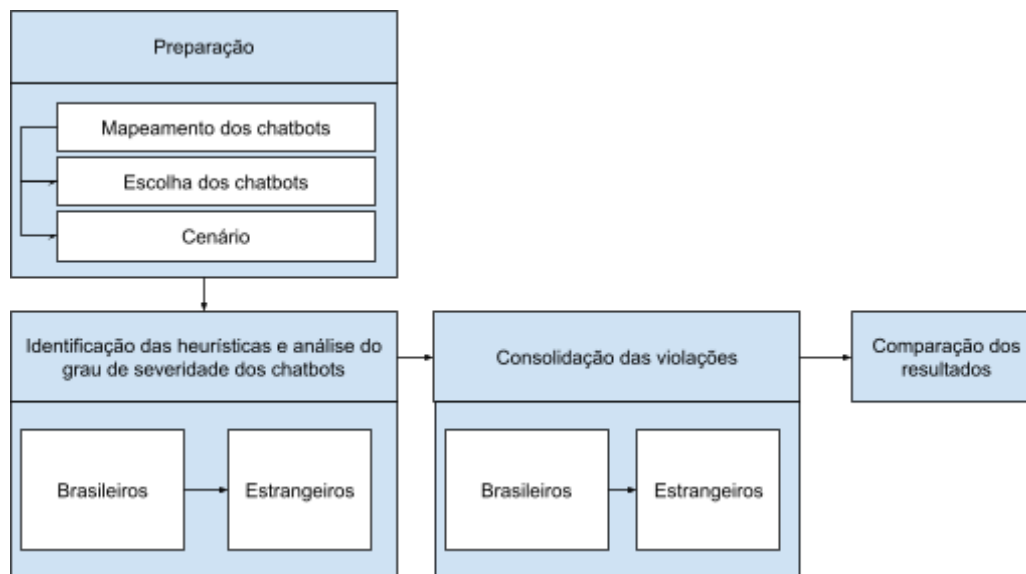
Para responder à questão de pesquisa “Quais são os potenciais impactos dos padrões de usabilidade violados em chatbots governamentais brasileiros e estrangeiros?” Foi utilizada uma versão adaptada das heurísticas de Nielsen [Nielsen, 1995; Nielsen e Molich, 1990] proposta por Langevin (2021). A avaliação heurística é um método amplamente difundido de avaliação de problemas de usabilidade [Bim, Salgado e Leitão, 2016].

A versão adaptada de Langevin (2021) contém as heurísticas tradicionais e duas específicas para chatbot. Com relação às tradicionais, algumas receberam ajustes para ser aplicadas ao contexto de avaliação de chatbots, outras mantiveram sua definição original proposta por Nielsen (Tabela 1).

**Tabela 1. Heurísticas de Langevin (2021)**

<b>Heurística</b>	<b>Descrição</b>
Visibilidade do estado do sistema	O chatbot sempre deve informar ao usuário sobre o status do sistema, dentro de um tempo razoável.
Correspondência entre o Sistema e o mundo real	O sistema deve entender e falar a linguagem dos usuários. Sem termos orientados ao sistema ou terminologias confusas.
Controle e Liberdade do usuário	O sistema deve dar suporte para o usuário desfazer e refazer comandos.
Consistência e Padrão	Os usuários não devem se perguntar se palavras, opções ou ações diferentes significam a mesma coisa. O chatbot deve ser capaz de enviar respostas consistentes, mesmo que comuniquem a mesma função de várias maneiras (e modalidades).
Prevenção de erros	O chatbot deve prevenir ou eliminar prontamente possíveis condições propensas a erros, verificando e confirmando com os usuários antes que eles executem uma ação.
Ajuda e orientação	O sistema deve orientar o usuário ao longo do diálogo, esclarecendo suas capacidades e ter os recursos de ajuda fáceis de recuperar e pesquisar.
Flexibilidade e eficiência de uso	O sistema deve ser eficiente tanto para especialista quanto para novatos, fornecendo aos usuários a modalidade e o hardware de entrada e saída apropriados (ou preferidos).
Design estético, minimalista e envolvente	O chatbot deve fornecer elementos de interação necessários para envolver o usuário e se adequar ao objetivo do sistema. Suportando interações curtas e expandindo a conversa, se o usuário escolher.
Ajude o usuário a reconhecer, diagnosticar e se recuperar de erros	O sistema deve indicar com precisão e clareza o problema (mensagens de erros), e sugerir uma solução de forma construtiva.
Preservação de contexto	O chatbot deve permitir ao usuário fazer referência a mensagens anteriores, fazendo relação entre os tópicos das sessões.
Confiabilidade	O sistema deve ser transparente quanto ao uso dos dados e de como o chatbot irá ser apresentado ao usuário.

O estudo foi conduzido por meio de 4 etapas e contou com a participação de três pesquisadores da área de Interação Humano-Computador (IHC), sendo dois com experiência de aplicação da Avaliação Heurística tradicional e que atuam em pesquisa com chatbots. A terceira pessoa pesquisadora é iniciante na área de IHC.



**Figura 1. Desenho do estudo**

Conforme apresentado na Figura 1, o estudo teve início na fase de preparação, dividida em três passos. No Mapeamento dos chatbots, em ambos os casos (chatbots brasileiros e estrangeiros), a busca pelos assistentes virtuais aconteceu por meio do mecanismo de busca da Google<sup>5</sup> de pesquisas em sites que continham informações sobre os *chatbots* ativos ou em publicações de trabalhos acadêmicos. Durante a busca, foram utilizados termos como *chatbot* (e sinônimos) acrescido de governo, governo eletrônico ou e-gov. A busca foi feita em português, inglês e espanhol.

Para selecionar os chatbots foram considerados critérios para a elegibilidade e seleção: I) ter o mínimo de interação sem a necessidade de inserir dados sigilosos, II) estar ativo e III) ser governamental ou prestador de serviço público. Depois de pesquisar e avaliar os *chatbots* brasileiros, o mesmo procedimento foi realizado para encontrar os *chatbots* governamentais de outros países que tivessem os mesmo critérios de elegibilidade já citados, acrescidos de IV) estar ativo para usuários fora do país de origem e V) ter funcionalidades parecidas com os *chatbots* brasileiros. Sendo o resultado de cada passo discutido e avaliado. Os *chatbots* que não atendessem os critérios citados acima são desconsiderados para esta pesquisa.

Para cada avaliação foi usado o cenário no qual os avaliadores interagiram com os assistentes com a finalidade de obter ajuda a determinada função. Por exemplo, no chatbot do RioCard<sup>6</sup>, um cidadão que busca ajuda para ajustes no cartão. No caso do Travelbot<sup>7</sup>, um cidadão buscando informações sobre a cidade que irá visitar no próximo semestre.

Na etapa de Identificação das heurísticas, todas as violações foram tabuladas e registradas com o auxílio de registros da interface que pudessem servir como evidências (sendo algumas apresentadas ao longo deste trabalho). Em seguida, a análise do grau

5 <https://www.google.com/>

6 <https://atende.riocardmais.com.br/>

7 <https://www.facebook.com/tfltravelbot>

severidade das violações. Neste método, a definição do grau envolve três fatores: I) Frequência com que o problema ocorre, II) Impacto do problema na interação dos usuários e III) Persistência do problema. As severidades são divididas em quatro graus de problemáticas: 1º Grau: Problemas Cosméticos não atrapalha a interação do usuário com sistemas; 2º Grau: Problema Pequeno, tem uma baixa prioridade em relação a reparos; 3º Grau: Problema Grande, esse tipo de problema prejudica fatores de usabilidade tidos como importantes para o projeto, alta prioridade de reparo; e 4º Grau: Problema Catastrófico, se mantido, provavelmente impedirá que o usuário realize suas tarefas e alcance seus objetivos. Na etapa de Consolidação, foi realizada a análise dos dados encontrados na etapa anterior de identificação das violações. Por fim, foi realizada uma comparação entre os chatbots brasileiros e estrangeiros.

## **4. Resultados**

Nesta seção serão apresentados os resultados da aplicação da avaliação heurística adaptada de Langevin (2021). Considerando a limitação de espaço, o foco dos resultados apresentados está nas heurísticas encontradas com maior grau de severidade e frequência. A seção está dividida em Mapeamento e seleção dos chatbots, Identificação das heurísticas violadas e análise do grau de severidade, Consolidação dos resultados e, por fim, a Comparação dos resultados encontrados nos chatbots brasileiros e estrangeiros.

### **4.1. Mapeamento e seleção dos chatbots**

Com relação aos chatbots brasileiros, ao todo foram encontrados 16, mas apenas 5 atenderam aos critérios de elegibilidade: Mauá, Sissa, Rose, SP Transp e Tomás. No caso dos chatbots de fora do país, ao todo, foram encontrados em torno de 12 (doze), porém apenas 5 foram selecionados, totalizando um de cada país: Emma (EUA), DGAE (Portugal), Boti (Argentina), Info Bot (Singapura) e Travel (Reino Unido).

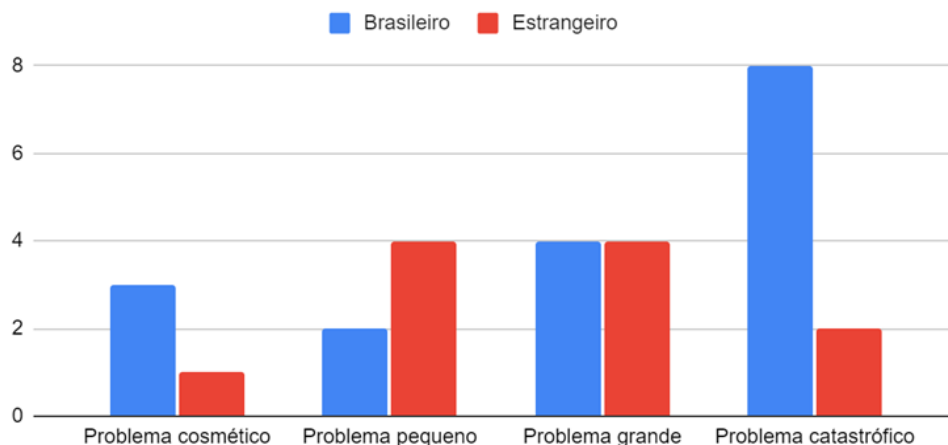
### **4.2 Identificação das heurísticas violadas nos chatbots e análise do grau de severidade**

Durante a avaliação dos dados coletados, *chatbots* estrangeiros e brasileiros apresentaram em média 2 a 3 violações durante as interações. Controle e Liberdade de Usuário apresentou um percentual de 21,4%, violação presente em 6 dos 10 *chatbots* estudados. A heurística Visibilidade do Status do Sistema teve 5 aparições e um percentual de 17,9%, seguido de Prevenção de Erros com 4 aparições, totalizando 14,3%. Em nenhuma das inspeções foram encontradas violações das heurísticas Correspondência entre o Sistema e o Mundo Real e Flexibilidade e Eficiência de uso.

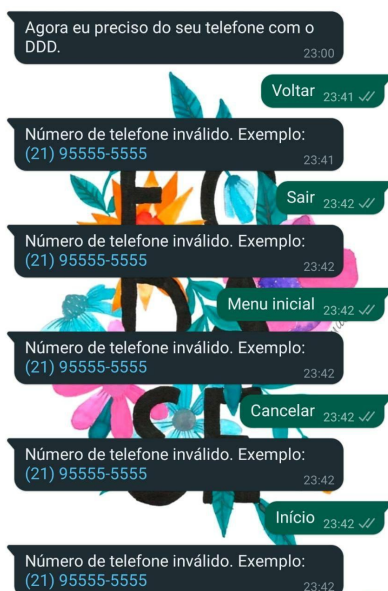
### **4.3. Consolidação das violações dos chatbots**

#### **4.3.1 Chatbots brasileiros**

Em interações de sistemas em funcionamento, ou seja, disponíveis para uso da população, a aparição do 4º Grau de severidade não deveria ser comum. Nota-se que há um alto índice de problemas catastróficos presentes nas interações dos assistentes virtuais brasileiros (Figura 2).



**Figura 2. Grau de severidade encontrado nos chatbots brasileiros e estrangeiros**



**Figura 3-a. Chatbot Tomás**



**Figura 3-b. Chatbot Boti**

O alto índice de violações catastróficas pode indicar que os chatbots provocam grandes problemas de usabilidade para as pessoas usuárias. Além disso, nos *chatbots* brasileiros, as heurísticas com maior índice de violações são Controle e Liberdade do Usuário, Confiabilidade e Ajude os usuários a reconhecer, diagnosticar e se recuperar de erros. Com relação a heurística Controle e Liberdade do Usuário, esta foi a heurística com o maior índice de violações encontradas no estudo, presente em 4 dos 5 *chatbots* brasileiros, todos com severidades de grau quatro. Sobre a heurística Ajude os usuários a reconhecer, diagnosticar e se recuperar de erros, também apresentou um alto índice. Na interação com o assistente Tomás (Imagem 3-a), por exemplo, o *chatbot* não informa e nem dá uma solução de saída para o usuário, tornando essa parte da interação um *loop*, a menos que seja encerrada a comunicação.

### 4.3.2 Chatbots estrangeiros

Nos *chatbots* estrangeiros as heurísticas com maiores violações são Visibilidade do status do sistema, Controle e Liberdade do Usuário e Prevenção de erros. Por exemplo, no chatbot Boti além de demorar, o *chatbot* não fornece a informação sobre “COVID em datas” (Figura 3-b). Apresentando uma violação da heurística de Visibilidade status do sistema.

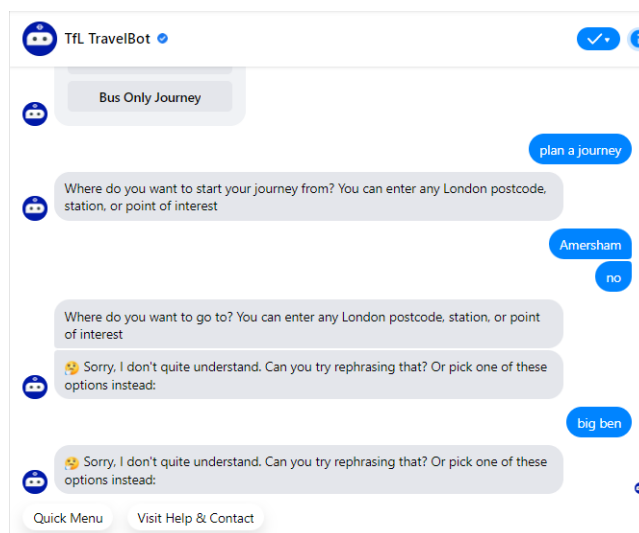


Figura 4. Chatbot Travelbot

A heurística Prevenção de erros trata da prevenção ou eliminação de possíveis condições propensas a erros e a verificação e confirmação com os usuários antes que eles executem uma ação. Foi o caso da interação com o Travel Bot (Figura 4). O *chatbot* solicitou o local de partida e chegada para criar a rota dos transportes disponíveis que o usuário poderia utilizar. Contudo, neste cenário, o usuário informou o local de partida ‘Amersham’. Quando o usuário tentou corrigir o erro e colocou o local correto ‘Big Ben’, o sistema não conseguiu compreender e a conversa foi interrompida.

### 4.4. Comparação entre chatbots brasileiros e estrangeiros

Dos *chatbots* brasileiros, apenas um não apresentou a violação de Controle e Liberdade do usuário e os que apresentaram tiveram sua severidade de grau catastrófico. Na Figura 5-a, é apresentada uma demonstração de como essa violação atrapalha o funcionamento do *chatbot* brasileiro RioCard. No qual, o usuário não consegue retornar para o menu anterior. O mesmo erro é encontrado no Assistente virtual da DGAE em Portugal (Figura 5-b).

A diferença entre os dois está na forma como o sistema resolve essa problemática. No caso do DGAE (Figura 5-b) o sistema permite que o usuário retorne utilizando a opção das FAQs, dessa forma o usuário consegue reverter o erro, fazendo com que a violação apresentasse um grau menor, o de Problema Pequeno. Já na interação do RioCard Mais (Figura 5-a), o sistema não entende o comando e impede que o usuário retorne e/ou saia da interação, forçando a pessoa usuária a iniciar uma nova conversa com o *chatbot*. Tornando esta violação com um grau maior, o de Problema Catastrófico.



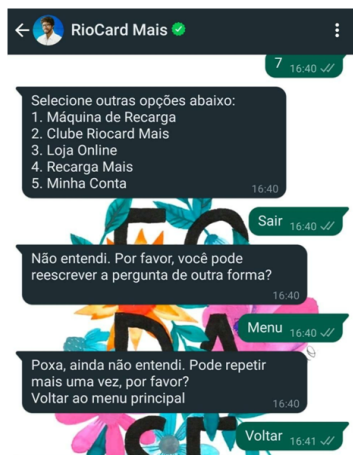


Figura 5-a. Chatbot RioCard Mais

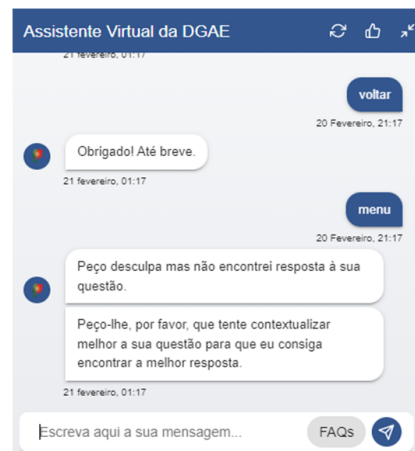


Figura 5-b. Chatbot DGAE

Permitir que o usuário desfça suas ações pode auxiliar que a pessoa usuária não necessite iniciar uma nova conversa para atingir o objetivo esperado, como no caso do chatbot RioCard Mais. Dependendo do domínio de aplicação, a quebra dessa heurística causa impactos não somente de interação e na forma como a pessoa usuária irá relacionar com a aplicação, mas para a pessoa em si, considerando um possível cenário de urgência. Dessa forma, o sistema deve estar preparado para cumprir e se preparar para desfazer a ação, caso seja solicitado.

Outro ponto importante é a questão da confiabilidade. Embora não tenha aparecido como uma das violações com maior recorrência entre as avaliações feitas, em alguns casos, a conversa entre cidadão e o chatbot prestador de serviços públicos, trata de dados sigilosos. Por exemplo, a Sissa, *chatbot* da Prefeitura de Assis (Figura 6-a), apresentou erro de grau 3 por não dar garantias sobre a privacidade dos dados solicitados. Por outro lado, o chatbot do RioCard Mais (Figura 6-b) não somente reconhece a importância do tratamento de dados sigilosos, mas também apresenta a política de privacidade.

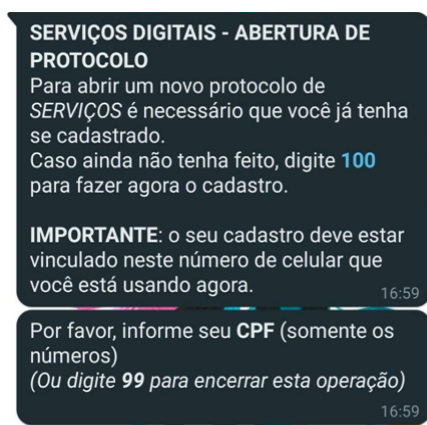


Figura 6-a. Chatbot Sissa

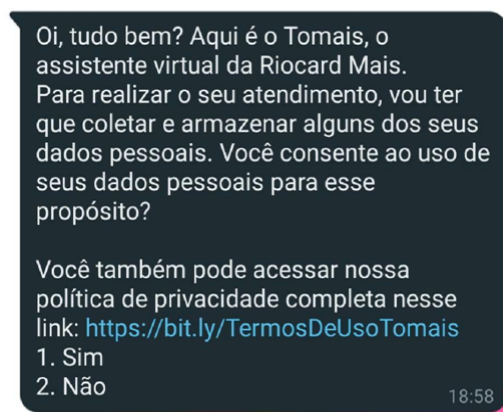


Figura 6-b. Chatbot RioCard Mais

Informações sigilosas como CPF, data de nascimento, telefone ou email, são constantemente solicitadas pelos agentes conversacionais governamentais. Desse modo, apresentar a política de segurança ou informar como esses dados irão ser utilizados estimula a confiança dos usuários sobre a aplicação. Quando não há esse tipo de interação presente nos *chatbots*, existe uma possibilidade maior da quebra na confiança do cidadão, pois os cidadãos que interagirem podem apresentar receio ao fornecer tais informações.

As violações dos *chatbots* estrangeiros também apresentaram falhas de confiabilidade, como por exemplo no caso do *chatbot* do governo de Singapura, que trata de assuntos da COVID 19. Na interação, não é apresentado no *chatbot* alguma mensagem informando que é um sistema automatizado, como é feito na interação do RioCard Mais. Mesmo que o nome do chatbot leve o nome *bot*, não existe uma linguagem clara e direta para o cidadão dizendo se o sistema é ou não automatizado. Falha considerada em grau de severidade Problema Pequeno. Essa falta de reconhecimento durante a interação pode causar dúvidas sobre quem está atendendo. Dúvida que pode surgir, principalmente, por usuários menos experientes. Nesse caso, o seu grau de severidade se torna baixo, pois não há uma quebra na interação e não causa danos na conversa.

## 5. Conclusões

Este trabalho teve como objetivo avaliar os problemas de usabilidade encontrados em chatbots de dentro e fora do país. Através da aplicação da avaliação heurística adaptada de Langevin (2021), foi possível encontrar as principais violações encontradas nos chatbots, respondendo à questão de pesquisa “Quais são os potenciais impactos dos padrões de usabilidade violados em chatbots governamentais brasileiros e estrangeiros?”.

Questões como controle e liberdade do usuário, visibilidade do status do sistema e prevenção de erros aparecem como os principais agravantes na relação entre o cidadão e a plataforma. Outro ponto importante a ser mencionado é a alta ocorrência de problemas catastróficos nos chatbots governamentais brasileiros. Neste sentido, é necessário uma atenção especial por parte dos projetistas dos chatbots voltados para prestação de serviços públicos. Fornecer uma boa experiência do usuário pode exigir uma avaliação da conversa, bem como das interações do usuário, de modo a transmitir com clareza os recursos da ferramenta.

Além disso, as violações encontradas nesta pesquisa, deixam claro quais são as heurísticas que devem ter uma maior atenção de projetistas. Em relação à severidade, não existe um padrão fixo para os graus, cada caso deve ser avaliado separadamente para entender qual é o problema e como ele afetará a adaptação dos usuários aos *chatbots*. Trabalhos futuros poderão ser dedicados a investigações e avaliações da linguística e comportamentos desses mesmo *chatbots* apresentados. Buscando compreender como o estilo de comunicação dos chatbots dentro e fora do Brasil podem apresentar resultados diferentes em termos de qualidade da interação e engajamento dos cidadãos. Uma vez que evidências preliminares indicam que chatbots fora do Brasil tendem a ter uma comunicação mais intimista com os usuários.

Por fim, perguntas sobre o impacto das falhas encontradas ainda ficam sem respostas. Necessitando uma avaliação com a participação dos usuários dos chatbots apresentados. Independentemente disso, a prestação com qualidade dos serviços públicos não deve ser a exceção, mas a regra. Ainda mais considerando que cada vez mais estes serviços podem ser oferecidos como uma alternativa a alta demanda populacional em meios eletrônicos.

### **Agradecimentos**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e financiado pelo CNPq (Processo No 139338/2021-7).

### **Referências**

- Bim, S. A., de Castro Salgado, e L. C., & Leitão, C. F. (2016). Evaluation by inspection: Comparing methods of practical, cognitive and semiotic basis. In Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems (pp. 1-10).
- Brandtzaeg P., e Følstad A. (2018). Chatbots. *Interactions* 25, 5 (Aug), 38–43. <https://doi.org/10.1145/3236669>.
- Carvalho, V., Cagnin, M., e Paiva, D. (2017). Avaliação de Acessibilidade de Web Sites de Governos Estaduais do Brasil. In Anais do XIII Simpósio Brasileiro de Sistemas de Informação (pp. 116-123). SBC.
- Dale, R. 2016. The return of the chatbots (2016). *Natural Language Engineering* 22, 811 – 817.
- Domingues, M. C. M., Melo, K. M. Miranda, A. F., de Freitas Cayres, A. Z., e 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 (pp. 251-258). SBC.
- Huai, J. (2011). “Quality Evaluation of E-Government Public Service,” in International Conference on Management and Service Science (MASS), August, Wuhan, pp. 1-4.
- Langevin, R. et al. Heuristic evaluation of conversational agents. (2021). In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021. p. 1-15.
- Monteiro, M., Vasconcelos, L., Viterbo, J., Salgado, L., e 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) (pp. 1178-1183). IEEE.
- Nielsen, J. (1995). How to conduct a heuristic evaluation. Nielsen Norman Group, 1(1), 8.

- Nielsen, J., e Molich, R. (1990). Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 249-256).
- de Oliveira, F. E. e 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 (pp. 125-125). SBC.
- Pereira, G. C. Monteiro, I. T. Vasconcelos, D. R., Braz, L., e Silva, C. , H. (2021). Classificação taxonômica de categorias de serviços públicos para aplicações digitais. In Anais do IX Workshop de Computação Aplicada em Governo Eletrônico (pp. 119-130). SBC.
- Quispe, F. E. M., e 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 (pp. 528-535). SBC.
- Santos, S. K. S., Mendes, N. F. O. e 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 (pp. 1-12). SBC.
- Siqueira, M., e Freire, A. (2020). Apresentação de denúncias sobre acessibilidade de serviços de governo eletrônico por pessoas com deficiência: uma análise à luz da teoria do comportamento planejado. In Anais Estendidos do XVI Simpósio Brasileiro de Sistemas de Informação (pp. 76-91). SBC.
- Teo, T., Srivastava, C., e Jiang, L. (2008). Trust and Electronic Government Success: An Empirical Study, *Journal of Management Information Systems*, 25:3, pp. 99-132, DOI: 10.2753/MIS0742- 1222250303
- Ventura I. Estudo mostra o perfil de reclamação dos chatbots. *Consumidor Moderno*. Encontrado em: <https://www.consumidormoderno.com.br/2021/11/25/queixas-comuns-desenvolve-chatbots/>. Acesso em: 20 de março de 2022.
- Zaidi, S., e M. K. Qteishat. (2012). “Assessing e-government service delivery (government to citizen),” *International journal of ebusiness and egovernment studies*, (4:1), 45-54.