

QUAL O LIMITE DA USABILIDADE? AVALIAÇÃO HEURÍSTICA DE *CHATBOTS* DAS ÁREAS DE SAÚDE E EDUCAÇÃO

EDUARDO ANTUNES BUENO¹
DAVI HENRIQUE GONCALVES TASSI²
PAULO ROBERTO DA SILVA RUIZ³

RESUMO

Este artigo analisa a usabilidade de *Chatbots* nas áreas de saúde e educação, com o objetivo de comparar a eficácia desses sistemas por meio de uma avaliação heurística de usabilidade. Na área de saúde, foram avaliados os dispositivos *ADA Health*, *Buoy Health*, *Your.MD* e *Dr. Wilson*. Já no setor educacional, foram analisados *Chatbots* de ensino de línguas: *Duolingo*, *Busuu*, *Babbel* e *Mondly*. Foi utilizada uma metodologia baseada em nove heurísticas sobre usabilidade, onde a partir da experiência de interação os dispositivos receberam notas classificando-os em quatro graus de problemáticas, sendo da mais simples a mais severa. As conclusões sugerem que, para melhorar a usabilidade, essas ferramentas precisam evoluir em termos de *design*, flexibilidade e adequação aos critérios de usabilidade, tanto para usuários novatos quanto experientes. O estudo destaca o potencial dos *Chatbots*, mas aponta desafios importantes que devem ser superados para maximizar sua eficácia em diferentes contextos.

Palavras-chave: Inteligência Artificial; Processamento de linguagem natural; Usabilidade.

ABSTRACT

This article analyzes the usability of *Chatbots* in the fields of healthcare and education, aiming to compare the effectiveness of these systems through a heuristic usability evaluation. In healthcare, the *ADA Health*, *Buoy Health*, *Your.MD*, and *Dr. Wilson* devices were assessed. In the educational sector, language learning *Chatbots*—*Duolingo*, *Busuu*, *Babbel*, and *Mondly*—were analyzed. A methodology based on nine usability heuristics was used, where interaction experiences were rated, classifying issues into four levels of severity, from minor to critical. The conclusions suggest that to improve usability, these tools need to evolve in terms of design, flexibility, and adherence to usability standards, both for novice and experienced users. The study highlights the potential of *Chatbots* but identifies significant challenges that must be overcome to maximize their effectiveness in different contexts.

¹Graduando, Ciência de Dados na Faculdade de Tecnologia de Adamantina – FATEC-Adamantina – Adamantina-SP. E-mail: eduardo.bueno2@fatec.sp.gov.br.

²Graduando, Ciência de Dados na Faculdade de Tecnologia de Adamantina – FATEC-Adamantina – Adamantina-SP.

³Docente, Faculdade de Tecnologia de Adamantina – FATEC-Adamantina – Adamantina-SP.

Key words: Artificial Intelligence; Natural Language Processing; Usability.

INTRODUÇÃO

Nos últimos anos, a utilização dos robôs de bate-papo, os populares *Chatbots*, apresentou um rápido crescimento. Esses sistemas automatizados, que utilizam Inteligência Artificial (IA) para interagir com os usuários, oferecem uma variedade de serviços que vão desde o suporte ao cliente até o auxílio em diversas atividades, incluindo rotinas de aprendizado (Zamora, 2017). A utilização desses dispositivos de conversação é uma realidade, diversos estudos e levantamento destacam a grande quantidade de pessoas que mantem contato com essa ferramenta e, na maioria dos casos, a aprovam devido ao fato de facilitar o acesso ao serviço ou atendimento que o usuário procura (Silveira, 2022).

Chatbots são *softwares* que simulam um serviço de conversa e permitem que os clientes façam perguntas como se estivessem conversando com uma pessoa. Essa tecnologia possui alta capacidade de processamento de informações, sendo configuradas e treinadas a partir de grandes bancos de dados para compreender e responder corretamente as mensagens dos usuários (Pereira; Dias, 2020). Essa tecnologia está inserida na área de Processamento de Linguagem Natural (PLN), a qual se caracteriza como uma forma de pesquisa que busca o processamento computacional da linguagem humana, que está em constante desenvolvimento. A palavra “natural” refere-se as línguas faladas em todo o mundo, diferenciada pelas linguagens matemáticas, de programação, gestuais dentre outras.

PLN está dividido em duas subáreas, sendo elas a Interpretação de Linguagem Natural - NLU (do inglês, *Natural Language Understanding*) e Geração de Linguagem Natural – NLG (do inglês, *Natural Language Generation*). NLU é a responsável pela análise (a compreensão das palavras segmentadas) e interpretação de uma língua (entender a escrita humana). Nos *Chatbots* a interpretação (NLU) é realizada ao receber a informação dos usuários, a escrita é analisada com base nos testes para a seguir ter a ação do sistema. Já a geração

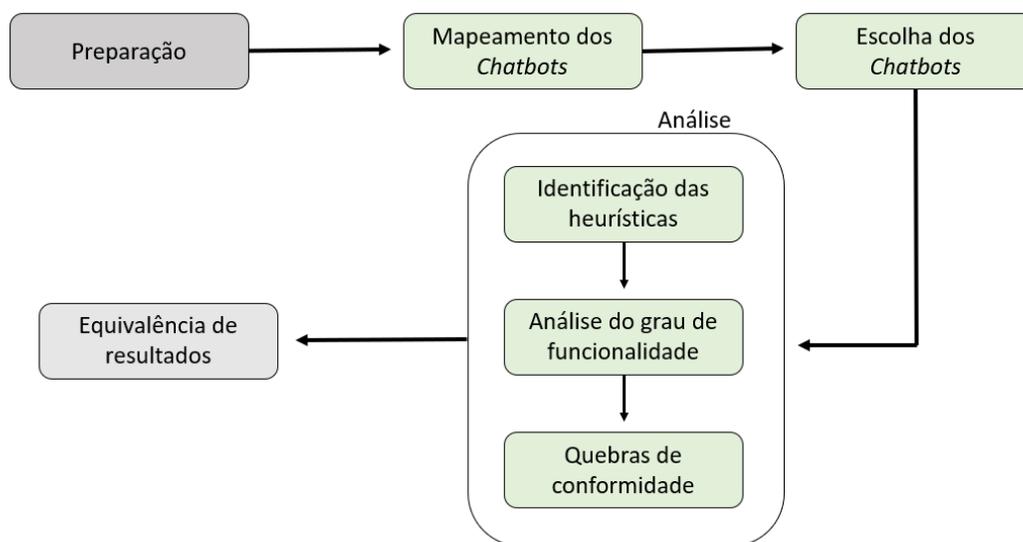
(*NLG*) está relacionada a obtenção de respostas, por exemplo o *ChatGPT*, uma IA que revolucionou o mercado por sua facilidade de entendimento e uma elevada taxa de acertos (Caseli; Nunes; Pagano, 2023).

Dada a relevância crescente dos *Chatbots* em diversos contextos e aplicações, este trabalho objetiva analisar a usabilidade e a eficácia desses sistemas no contexto de serviços de saúde e educação. A pesquisa aborda as funcionalidades específicas dos dispositivos em cada área, explorando como essas tecnologias estão sendo implementadas e avaliando seu impacto na experiência dos usuários. Para isso, foi utilizada uma metodologia baseada em estudos de caso e análise comparativa, visando identificar os principais desafios e oportunidades na aplicação dos *Chatbots*.

MATERIAL E MÉTODOS

Este trabalho foi desenvolvido seguindo as etapas apresentadas no fluxograma metodológico (Figura 1), descritas na sequência.

Figura 1. Fluxograma metodológico.



Fonte: Adaptado de Batista, Monteiro e Salgado (2022).

A etapa de preparação envolveu a organização e planejamento a fim de compreender os conceitos, características e tipos de *Chatbots* no contexto do PLN. A etapa de mapeamento envolve a identificação e o levantamento dos *Chatbots*, escolhidos para a análise. Foram definidas áreas prioritárias para serem exploradas, sendo elas a saúde e a educação, já que há um crescimento significativo de utilização dessa tecnologia nessas áreas (Araújo, Bessa, Faria, 2023; Cardoso *et al.*, 2023).

Após o mapeamento, ocorre a seleção dos *Chatbots*, objeto deste estudo. Os critérios definidos para a escolha foram sua utilização, o contexto no qual está inserido na área, sua abrangência e tipo específico de dispositivo, referindo-se a sua característica de utilização. Com isso, foram selecionados os seguintes *Chatbots*:

- *ADA Health*: Auxilia no diagnóstico de saúde com base nos sintomas fornecidos pelo usuário, oferecendo sugestões de acordo com os sintomas descritos e apresenta um relatório ao final da consulta. É objetivo, sem uso de elementos visuais como emojis, e tem uma interface simples e funcional.
- *Buoy Health*: Proporciona ajuda no diagnóstico, mas oferece opções mais limitadas em relação aos sintomas e diagnósticos. Embora útil, é considerado simplista e pode não fornecer diagnósticos mais detalhados. Sua linguagem é formal e objetiva, sem uso de componentes visuais.
- *Your.MD*: Permite ao usuário inserir sintomas e fornece sugestões baseadas nessas informações. Ele é fácil de usar e corrige erros ortográficos, mas oferece opções de diagnóstico limitadas. Assim como os outros, utiliza uma linguagem formal e direta.
- *Dr. Wilson*: Responde diretamente com base no sintoma principal fornecido. É simples e direto, focado em fornecer informações básicas, mas com pouca flexibilidade. Não oferece muitos elementos visuais e permite acesso contínuo à consulta durante a interação.
- *Duolingo*: Utilizado no aprendizado de idiomas, é interativo e disponibiliza exercícios práticos. Oferece *feedback* imediato com base nas respostas dos

usuários e mantém um cronograma de atividades, sem permitir pular etapas. Utiliza linguagem formal e oferece estatísticas de desempenho ao final das lições.

- *Busuu*: Trata-se de uma plataforma de aprendizado de idiomas. Oferece respostas baseadas em seleção e mantém o usuário dentro de um cronograma específico de aprendizagem. Não permite muitas opções de navegação ou sinônimos, mas fornece *feedback* claro sobre erros e solicita confirmações durante o uso.

- *Babbel*: Focado no ensino de idiomas por meio de atividades de seleção de respostas. Ele é simples e direto, com um *design* minimalista. Não oferece muita flexibilidade de uso ou opções de correção de erros ortográficos, mas permite refazer atividades se o usuário cometer erros.

- *Mondly*: Focado no aprendizado de idiomas, onde o usuário interage principalmente por meio de seleção de alternativas. Ele oferece uma interface simples e funcional, com *feedback* imediato e um *design* minimalista.

Para a consecução da análise, primeiramente foram definidas as heurísticas adotadas na avaliação. Os procedimentos aqui adotados foram adaptados do trabalho de Batista, Monteiro e Salgado (2022), no qual os autores realizaram uma avaliação de *Chatbots* governamentais. Nesse sentido, foi utilizada uma versão adaptada das heurísticas de Nielsen (Nielsen, 1995) proposta por Langevin *et al.* (2021). Essa versão, possui as heurísticas tradicionais e duas específicas para *Chatbots* (Tabela 1).

Na análise foram contempladas as etapas de avaliação de funcionalidade dos *Chatbots* conforme as heurísticas definidas. A seguir, na etapa de concentração de quebras de conformidade, é analisado onde e como os dispositivos não atendem às expectativas ou requisitos previamente estabelecidos. A "quebra de conformidade" se refere às áreas em que os *Chatbots* falham em seguir as heurísticas ou padrões esperados. Essa avaliação foi realizada por meio do grau de severidade das violações, envolvendo três fatores: 1) Frequência com que o problema ocorre; 2) Impacto do problema na interação dos usuários e 3) Persistência do problema. As

Qual o limite da usabilidade? Avaliação heurística de <i>Chatbots</i> das áreas de saúde e educação.	Eduardo A. Bueno; Davi Henrique G. Tassi; Paulo R. da Silva Ruiz.
--	---

severidades são divididas em quatro graus, distribuídos em notas de 0 a 10: 1º Grau (6 a 10): problemas não atrapalham a interação do usuário com o sistema; 2º Grau (4 a 6): pequenos problemas com baixa prioridade de reparos; 3º Grau (2 a 4): grandes problemas que prejudicam fatores de usabilidade, possuindo alta prioridade de reparo; e 4º Grau (0 a 2): problema catastrófico, que se mantido impedirá que o usuário realize suas tarefas e alcance seus objetivos.

Tabela 1. Heurística utilizadas.

HEURÍSTICA	DESCRIÇÃO
1. Visibilidade do estado do sistema.	O <i>Chatbot</i> sempre deve informar ao usuário sobre o status do sistema, dentro de um tempo razoável.
2. Correspondência com o mundo real.	O sistema deve entender e falar a linguagem dos usuários. Sem termos orientados ao sistema ou terminologias confusas.
3. Controle do usuário e liberdade.	O sistema deve dar suporte para o usuário desfazer e refazer comandos.
4. Consistência e padrões.	Os usuários não devem se perguntar se palavras, opções ou ações diferentes significam a mesma coisa. O <i>Chatbot</i> deve ser capaz de enviar respostas consistentes, mesmo comunicando a mesma função de várias maneiras e modalidades.
5. Prevenção de erros.	O <i>Chatbot</i> deve prevenir ou eliminar prontamente possíveis condições propensas a erros, verificando e confirmando com os usuários antes que eles executem uma ação.
6. Reconhecimento x Memória.	Opções claras por meio de visual descritivo com elementos e instruções explícitas. Mostra o resumo das informações coletadas antes das transações, oferecendo menu e ajuda.
7. Flexibilidade e eficiência de uso.	O sistema deve ser eficiente tanto para especialista quanto para novatos, fornecendo aos usuários a modalidade e o <i>hardware</i> de entrada e saída apropriados ou preferidos.
8. Estética e minimalismo.	O <i>Chatbot</i> 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.
9. Ajuda, diagnóstico e recuperação de erros.	O sistema deve indicar com precisão e clareza o problema (mensagens de erros), e sugerir uma solução de forma construtiva.

Fonte: Adaptado de Langevin et al., (2021).

Por fim, os resultados da análise são comparados para determinar a consistência ou equivalência entre os *Chatbots* avaliados, incluindo uma análise de como se comparam entre si em termos de desempenho, funcionalidade e aderência às heurísticas.

Chatbots

Os primeiros dispositivos surgiram na década de 1960, como o “*Eliza*”, criado por Joseph Weizenbaum, que simulava uma conversa terapêutica com os pacientes. Com o tempo outros *Chatbots* foram desenvolvidos, como o “*Parry*” e o “*ALICE*”, melhorando a capacidade de manter diálogos complexos, contando com treinamento em bases de dados mais robustas, as quais proporcionaram avanços significativos nessa área (Moraes; De Souza, 2015).

Atualmente, com o avanço da IA e do Aprendizado de Máquina (AM), essa tecnologia sofreu transformações importantes, sendo capaz de compreender comandos complexos e manter diálogos naturais com seus usuários. São utilizados em diversas plataformas, como redes sociais e comércio eletrônico, oferecendo uma experiência personalizada e eficiente (Martins; Carvalho, 2021).

Essa tecnologia é ótima para ajudar os usuários a encontrar informações rapidamente. Eles respondem perguntas e solicitações por texto, áudio e arquivos, sem precisar de pesquisa manual. Atualmente, essas ferramentas podem ser encontradas em alto-falantes inteligentes residenciais, em serviços de mensagens como SMS, *WhatsApp* e *Facebook Messenger*, e até em aplicativos de trabalho como *Slack* (García-Reina, 2018).

Os *Chatbots* mais avançados, conhecidos como "assistentes virtuais inteligentes" ou "agentes virtuais", usam tecnologia de ponta para entender conversas mais naturais e automatizar tarefas. Além de assistentes como *Siri*, *Alexa*, *Google Gemini* e *ChatGPT*, esses agentes também estão se tornando comuns em empresas, ajudando tanto clientes quanto funcionários (Martins; Carvalho, 2021). A IA é uma ferramenta poderosa para estratégias de *marketing*. *Chatbots* de IA podem

oferecer suporte ao cliente, entender seu comportamento e preferências e criar interações personalizadas (Silva; Almeida, 2019).

RESULTADOS E DISCUSSÃO

O processo utilizado pelos *Chatbots* da área de saúde analisados neste trabalho é guiado e informativo. Como exemplo, na Figura 2 o *ADA Health* coleta dados dos usuários de forma precisa e rápida, questionando se está com febre, o que ajuda a iniciar a avaliação dos sintomas. A partir da resposta, o dispositivo pergunta sobre o sintoma de tosse, oferecendo informações sobre possíveis causas, como resfriado comum, COVID-19 e laringite aguda. As perguntas se aprofundam dependendo dos sintomas fornecidos pelo usuário. Por fim, o assistente fornece um possível diagnóstico baseado em seu banco de dados a partir de casos semelhantes. O *Buoy Health* segue o mesmo processo do *ADA Health*, mas com perguntas menos informativas, o que gera dúvidas para o usuário. O *Your.MD* também segue a mesma lógica, mas peca nas poucas opções de resultados. Por fim, o *Dr. Wilson* não segue essa lógica, diferenciando-se por ser mais informativo em relação a cuidados de saúde, sendo treinado não para diagnosticar, mas para orientar a procurar um especialista.

Figura 2. Interface do *Chatbot ADA Health*.

Nova avaliação × **Tosse** ... **Tosse** ...

^ Anterior

Pessoas com sintomas semelhantes aos seus geralmente podem fazer a gestão deles com segurança em casa. Você também pode procurar aconselhamento visitando ou contactando sua farmácia local. Se seus sintomas persistirem por mais tempo do que o esperado, se eles piorarem, ou se você notar novos sintomas, deve procurar uma avaliação médica adicional.

Causas possíveis

1 **Resfriado comum**
 Geralmente pode ser controlado em casa
 6 entre 10 pessoas com esses sintomas têm essa condição.
 Mostrar detalhes

2 **Doença por coronavírus 2019 (COVID-19)**
 Geralmente pode ser controlado em casa
 2 entre 100 pessoas com esses sintomas têm essa condição.
 Mostrar detalhes

3 **Laringite aguda**
 Geralmente pode ser controlado em casa

Causas possíveis

1 **Sinusite bacteriana**
 Procure assistência médica
 4 entre 10 pessoas com esses sintomas têm essa condição.
 Mostrar detalhes

2 **Febre do feno**
 Procure assistência médica
 1 entre 10 pessoas com esses sintomas têm essa condição.
 Mostrar detalhes

3 **Sinusite viral**
 Procure assistência médica
 1 entre 10 pessoas com esses sintomas têm essa condição.
 Mostrar detalhes

4 **Rinite não-alérgica**
 Procure assistência médica
 9 entre 100 pessoas com esses sintomas têm essa condição.
 Continuar

Você tem febre?
 O que significa isso?

Sim
 Não
 Eu não sei

Dê sua opinião

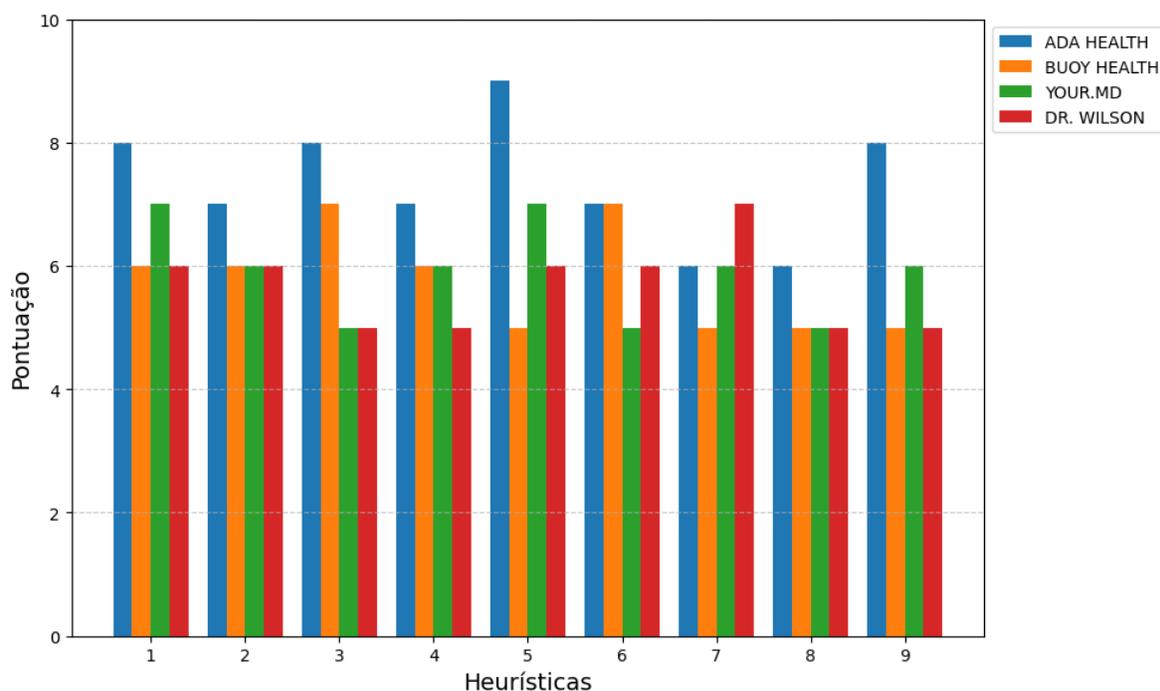
Continuar

Fonte: Elaborado pelos autores, (2024).

Considerando a quantidade de heurísticas avaliadas no 1º Grau, o *Ada Health* foi aquele que apresentou o melhor desempenho, seguido do *Your.MD* (Figura 3). Já o *Dr. Wilson* apresentou o pior desempenho. Não houve heurística classificada nos 3º e 4º Graus. Nas heurísticas 1 e 2 (visibilidade de estado do sistema e correspondência com o mundo real) todos foram classificados no 1º Grau. Na heurística 3 (controle do usuário e liberdade) o *Your.MD* e *Dr. Wilson* foram classificados no 2º grau, com baixa prioridade de reparos. Considerando a heurística 4 (consistência e padrões) apenas o *Dr. Wilson* foi classificado no 2º grau. Por sua vez, as heurísticas 5 (prevenção de erros) e 7 (flexibilidade e eficiência de uso) revelaram o *Buoy Health* necessitando de baixa prioridade de reparos. Somente o *Your.MD* necessita de pequenos reparos relacionados a reconhecimento e memória

(H6). Quanto a estética e minimalismo (H8) apenas o *Ada Health* não necessita de pequenos reparos. Por fim, a heurística 9 (ajuda, diagnóstico e recuperação de erros) revela a necessidade de pequenos reparos nos *Chatbots Buoy Health* e *Dr. Wilson*.

Figura 3. Desempenho dos *Chatbots* da área de saúde.



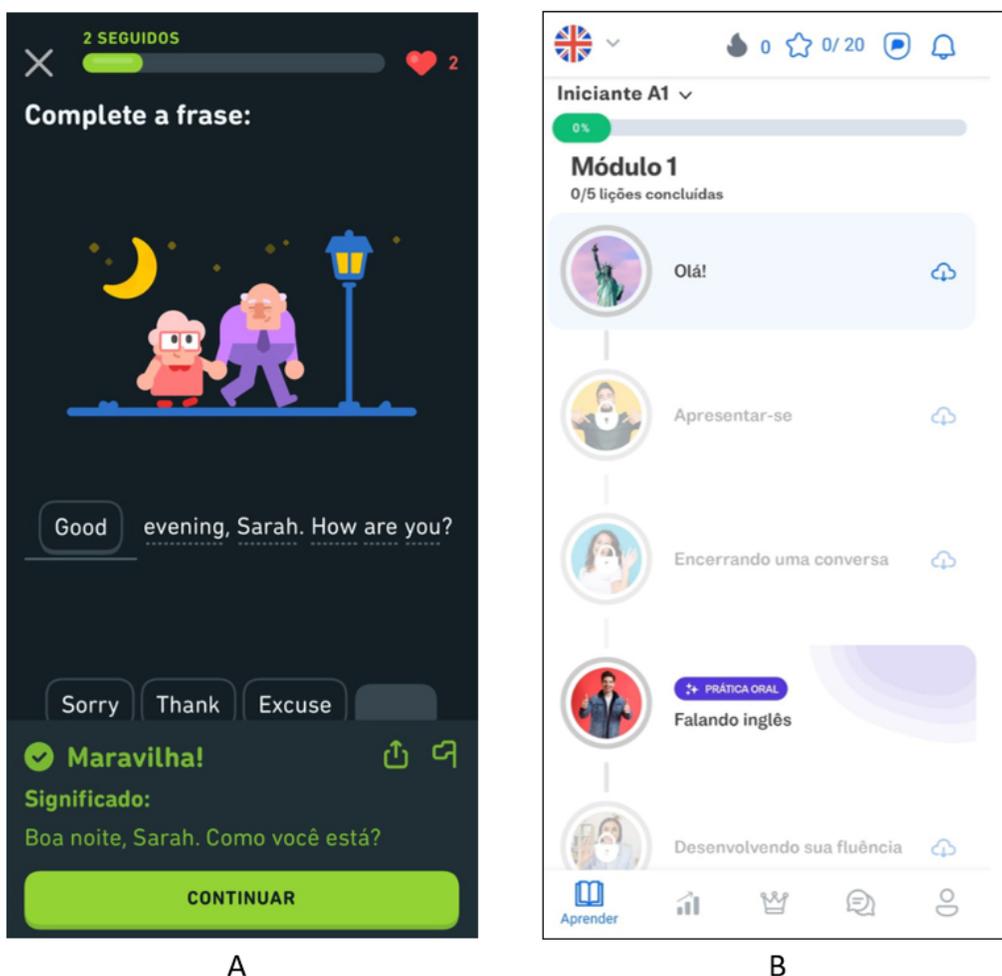
Fonte: Elaborado pelos autores, (2024).

Os *Chatbots* de saúde apresentam características comuns que refletem sua utilidade como ferramentas de triagem inicial para usuários que buscam orientações rápidas e diretas sobre sintomas. No entanto, embora todos cumpram suas funções básicas, existem limitações em termos de personalização, profundidade de diagnóstico e interação visual. O *ADA Health*, por exemplo, oferece um processo eficiente e orientado por seleção de sintomas, mas carece de um *design* mais sofisticado e maior flexibilidade nas respostas. O *Buoy Health*, apesar de ser funcional, apresenta limitações nas opções de diagnóstico, o que pode frustrar usuários que buscam uma abordagem mais abrangente. *Your.MD* se diferencia ao permitir maior interatividade com o usuário por meio da confirmação de sintomas,

mas, assim como os demais, peca pela simplicidade e pela ausência de um sistema mais robusto de navegação. Já o *Dr. Wilson*, apesar de ser informativo e direto, também não oferece uma interface interativa que permita refazer ações ou acessar consultas anteriores de forma prática.

Os *Chatbots* educacionais da área de idiomas oferecem uma plataforma interativa em que o usuário segue um cronograma de estudo (Figuras 4 e 5). Todos eles utilizam comandos escritos não sendo tolerantes a erros ortográficos. Além disso, fornecem *feedback* de diferentes formas, com estatísticas de desempenho ao final de cada etapa do aprendizado.

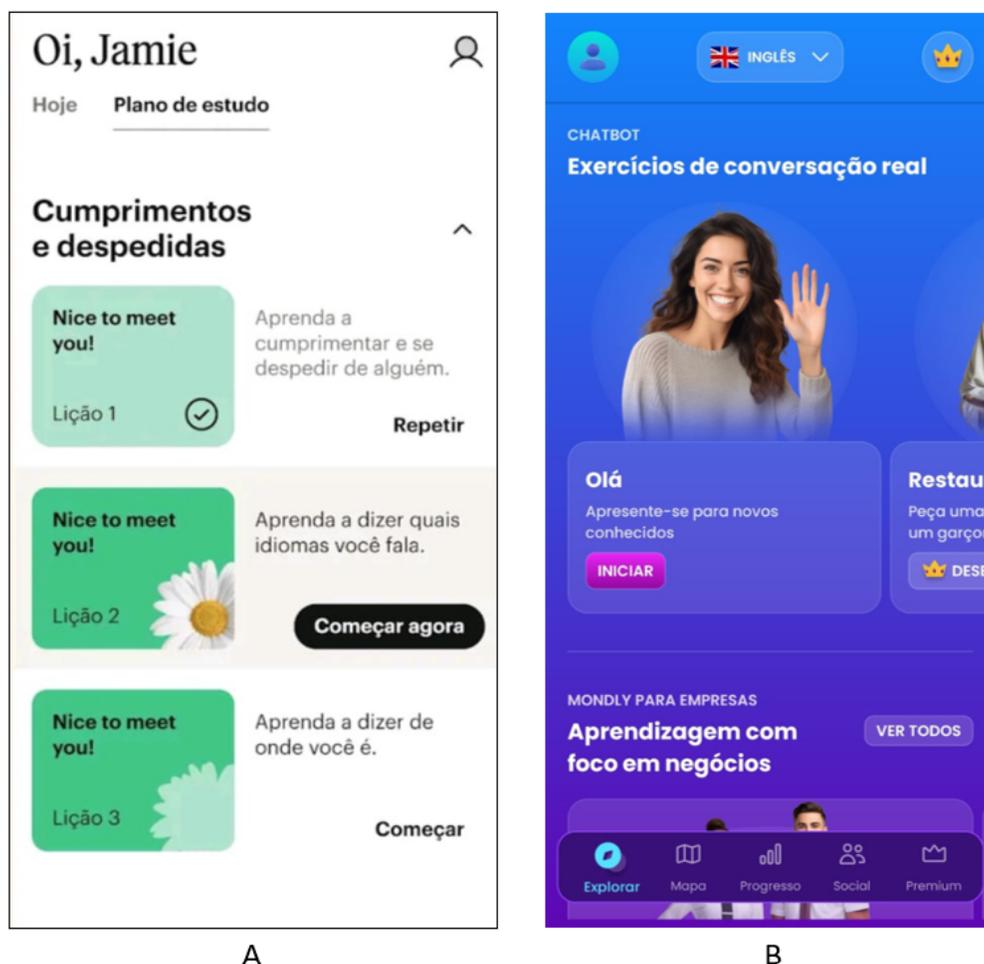
Figura 4. Interface dos *Chatbots*: (A) *Duolingo*, (B) *Busuu*.



A

B

Fonte: Elaborado pelos autores, (2024).

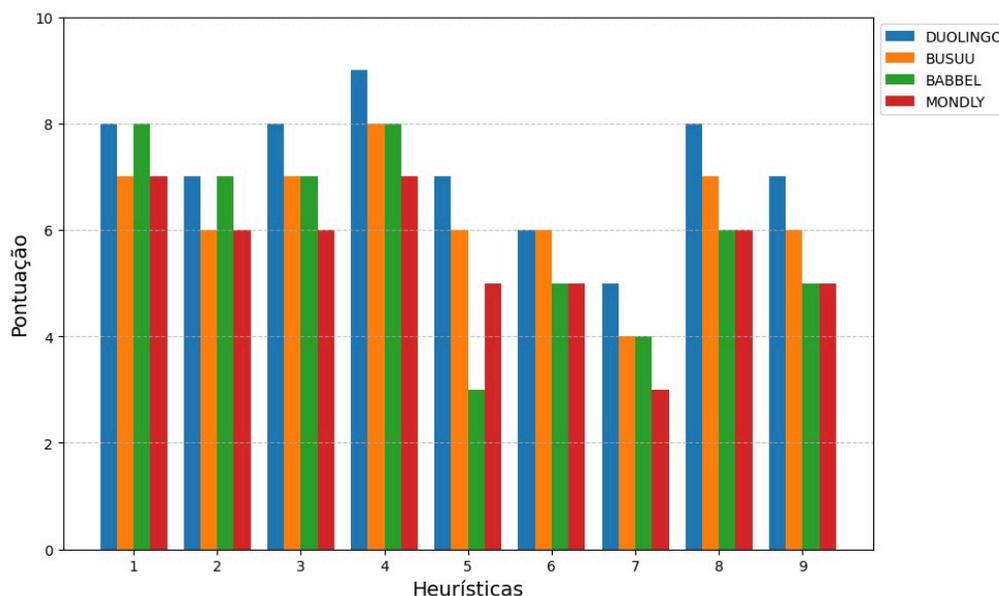
Figura 5. Interface dos *Chatbots*: (A) *Babbel*, (B) *Mondly*.

A

B

Fonte: Elaborado pelos autores, (2024).

O *Duolingo* e o *Busuu* apresentaram os melhores resultados, com 8 heurísticas classificadas como 1º Grau (Figura 6). Nas heurísticas 6 (reconhecimento x memória) e 9 (ajuda, diagnóstico e recuperação de erros) o *Babell* e *Mondly* foram classificados no 2º Grau, revelando baixa necessidade de reparos. Em prevenção de erros (H5) o *Mondly* necessita de baixos reparos, já o *Babell* foi classificado no 3º Grau, alertando para alta prioridade de reparos. Quanto a heurística 7 (flexibilidade e eficiência de uso) todos os dispositivos necessitam de reparos, sendo que o *Mondly* de alta e os demais de baixa prioridade. Por fim, nas demais heurísticas todos os *Chatbots* foram classificados no 1º Grau, onde os pequenos problemas identificados não atrapalham a interação do usuário com o sistema.

Figura 6. Desempenho dos *Chatbots* da área de educação.

Fonte: Elaborado pelos autores, (2024).

Em termos gerais, os *Chatbots* avaliados das áreas de saúde e educação oferecem respostas concisas e visam eficiência no uso, mas carecem de personalização avançada e recursos de navegação. A ausência de uma estética visual mais sofisticada ou de elementos de interface que melhorem a usabilidade é um ponto a ser aprimorado. Apesar dessas limitações, eles demonstram potencial para servir como ferramentas auxiliares, no caso da saúde na triagem inicial e orientação básica, já na educação como aprofundamento de estudos. Os resultados das análises das heurísticas demonstram que a interação promove o alcance dos objetivos aos quais os dispositivos se propõem a oferecer. Diferente dos resultados de Batista, Monteiro e Salgado (2022), onde os autores identificaram violações graves na usabilidade de *Chatbots* governamentais, os dispositivos aqui analisados apresentaram pequenas violações, restritas a heurísticas específicas. Na área de saúde os resultados foram mais homogêneos quando comparados com a área da educação, na qual os dispositivos apresentaram mais problemas nas heurísticas 5 (prevenção de erros), 6 (reconhecimento e memória) e 7 (flexibilidade e eficiência de uso), onde apenas dois casos necessitam de alta prioridade de reparos.

CONCLUSÃO

Este trabalho teve como objetivo avaliar os problemas de usabilidade encontrados em *Chatbots* das áreas de saúde e educação. Através da aplicação da avaliação heurística adaptada de Langevin (2021), foi possível encontrar as principais violações verificadas nos dispositivos. A análise comparativa dos *Chatbot* revelou que, apesar de desempenharem funções básicas de triagem e suporte, essas ferramentas ainda apresentam limitações, especialmente em termos de personalização e profundidade nas interações. O *ADA Health* e o *Buoy Health*, fornecem diagnósticos rápidos, mas carecem de flexibilidade e design sofisticado. No setor educacional, ferramentas como *Duolingo* e *Babbel* mantêm uma estrutura rígida, oferecendo *feedback* imediato, mas limitando a navegação e a adaptação às necessidades do usuário. Assim, conclui-se que, embora eficazes em suas áreas específicas, os *Chatbots* precisam evoluir para fornecer uma experiência mais imersiva e personalizada, especialmente em termos de interface e adequação às heurísticas de usabilidade. E, embora sejam boas ferramentas para diagnóstico de doenças e aprofundamento de estudos, elas não substituem os profissionais que atuam nessas áreas, sendo apenas ferramentas auxiliares em suas respectivas áreas.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, E.; BESSA, L.; FARIA, A. C. A potencialização do uso dos *Chatbots* em atendimentos na área da saúde a partir da pandemia da COVID-19: um estudo biométrico (Sistemas de Informação). **Repositório Institucional**, v. 1, n. 1, 2023.

BATISTA, G. O. S.; MONTEIRO, M. S.; SALGADO, L. C. C. Investigando *Chatbots* governamentais: um panorama sobre a usabilidade dentro e fora do Brasil. In: **Anais do X Workshop de Computação Aplicada em Governo Eletrônico**. SBC, 2022. p. 73-84.

CARDOSO, F. S. et al. O uso da inteligência artificial na educação e seus benefícios: uma revisão exploratória e bibliográfica. **Revista Ciência em Evidência**, v. 4, n. FC, p. e023002-e023002, 2023.

CASELI, H.; NUNES, M. G. V.; PAGANO, A. Understanding natural language processing. **Journal of Computational Linguistics**, v. 12, n. 3, p. 45-67, 2023.

GARCÍA-REINA, L. F. **Asistente virtual de tipo ChatBot**. Trabalho de Graduação. Universidad Católica de Colombia. Facultad de Ingeniería. Programa de Ingeniería de Sistemas. Bogotá, Colombia, 2018.

LANGEVIN, R. et al. Heuristic evaluation of conversational agents. In: **Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems**, 2021. p. 1-15.

MARTINS, J. R.; CARVALHO, P. D. Avanços em processamento de linguagem natural aplicados a Chatbots. **Revista Brasileira de Inteligência Artificial**, v. 7, n. 1, p. 21-34, 2021.

MORAES, S. M. W; DE SOUZA, L. S. Uma abordagem semiautomática para expansão e enriquecimento linguístico de bases de chatbots. In: **Congresso Internacional de Informática Educativa**, 2015. p. 600-605.

NIELSEN, J. **How to conduct a heuristic evaluation**. Nielsen Norman Group, v. 1, n. 1, p. 8, 1995.

PEREIRA, M. R.; DIAS, L. S. Chatbots como ferramenta de suporte ao atendimento ao cliente. **Revista Brasileira de Tecnologia da Informação**, v. 9, n. 2, p. 55-68, 2020.

SILVA, A. P.; ALMEIDA, F. O. Uso de chatbots na educação: potencialidades e desafios. **Revista de Educação e Tecnologia**, v. 14, n. 3, p. 89-104, 2019.

SILVEIRA, A. L. R. **Confiança do usuário em chatbots: uma revisão bibliográfica**. Trabalho de Conclusão de Curso (Graduação em Sistemas e Mídias Digitais) — Universidade Federal do Ceará, Fortaleza, 2022.

ZAMORA, J. I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot perception and expectations. In: **Proceedings of the 5th International Conference on Human Agent Interaction**, Bielefeld, Germany, p. 253–260, out. 2017.