

# ARIA e Interactive Access: projetando chatbots para idosos

Carlos Nery Ribeiro, Cynthia Letícia Teles de Oliveira, Lucas Padilha Modesto de Araujo,  
Kamila Rios da Hora Rodrigues, Marcelo Garcia Manzato  
carlosnribeiro@usp.br, cynthya@usp.br, padilha.lucas@usp.br, kamila.rios@icmc.usp.br, mmanzato@icmc.usp.br  
Instituto de Ciências Matemáticas e de Computação  
Universidade de São Paulo  
São Carlos, SP

## ABSTRACT

Elderly individuals represent a significant portion of the population, and trends indicate that this number will increase at an accelerating rate. The elderly often face challenges when interacting with conversational systems, whether due to a lack of familiarity or age-related limitations, such as reduced visual acuity and dexterity. Therefore, it is essential to consider the needs of the elderly when designing these systems. In this work, we present the Assistente de Recomendação e Interação Acessível (ARIA), an accessible chatbot for the elderly that provides recommendations for cultural content, and Interactive Access, a set of guidelines for building accessible chatbots for the elderly. Performance and usability tests were conducted on both artifacts, based on Nielsen's heuristics. The artifacts showed positive results, pointing the next steps for studies with elderly users.

## KEYWORDS

Sistema conversacional, grandes modelos de linguagem, diretrizes, usabilidade, idosos.

## 1 INTRODUÇÃO

Uma pessoa é legalmente considerada idosa a partir dos 60 anos. De acordo com o censo realizado em 2022 pelo IBGE, o Brasil possui mais de 32 milhões de pessoas idosas, o que corresponde a quase 16% da população brasileira. A população idosa do Brasil dobrou desde o último censo em 2010, o que corrobora com o envelhecimento da população [5].

Ao mesmo tempo, também se nota o avanço da tecnologia em diversos aspectos e com capacidade de promover diferentes tipos de suporte aos usuários. Entretanto, quando verificamos o uso de ferramentas tecnológicas por pessoas idosas se nota a baixa aderência dos mesmos. Isso ocorre porque muitos idosos foram apresentados ao computador e *smartphone* em fase senil e têm limitações motoras causadas pela idade. Assim, se faz necessário que as interfaces e interações sejam pensadas para o público idoso a fim de incluí-lo digitalmente [10].

Para que sejam desenvolvidas soluções para idosos, é necessário que as pessoas responsáveis pelos projetos, bem como quem os desenvolve, possuam vasto conhecimento em usabilidade e acessibilidade, o que nem sempre é verdade. Atualmente há algumas organizações que definem padrões e diretrizes que, inclusive, são

bem consolidados na literatura, porém a forma em que são apresentadas dificulta o acesso e o entendimento por parte de quem deseja conhecer.

Pensando nisso, este trabalho procurou tornar as diretrizes produzidas em uma etapa anterior deste projeto, conforme detalhado na dissertação de mestrado da Universidade de São Paulo [1], mais intuitivas, bem como disponibilizá-las no Notion<sup>1</sup>, permitindo acesso fácil e gratuito. Além disso, também foi produzido o Assistente de Recomendação e Interação Acessível (ARIA), um *chatbot* baseado em LLM (*Large Language Model*) que faz recomendações de conteúdos culturais e é acessível para idosos. Esses dois produtos da pesquisa foram avaliados por especialistas e apresentaram resultados positivos.

Este trabalho está organizado da seguinte maneira: a Seção 2 traz uma visão geral dos trabalhos relacionados; na Seção 3 é apresentada a metodologia aplicada no desenvolvimento dos artefatos; os estudos realizados para avaliar os artefatos estão na Seção 4; a Seção 5 apresenta os resultados obtidos; e a Seção 6 traz as considerações finais.

## 2 TRABALHOS RELACIONADOS

Chen et al. [2] realizaram uma análise bibliométrica sobre a interação de humanos com sistemas conversacionais, abrangendo 3013 publicações de 2000 a 2022. O estudo identificou tendências principais, como design de *chatbots*, experiência do usuário, expressão emocional, privacidade e ética. Os resultados mostram um crescimento exponencial de publicações nos últimos cinco anos, com foco crescente na experiência do usuário e comunicação emocional.

Furini et al. [4] exploraram o impacto das interfaces de conversação, como *chatbots*, na acessibilidade web. O artigo discute os benefícios dessas interfaces para pessoas com deficiência, como as visuais e motoras, permitindo o acesso à informação através da interação por voz. Os autores concluem que, embora as interfaces de conversação ofereçam benefícios significativos para a acessibilidade, ainda há problemas a serem resolvidos para garantir uma experiência de usuário ideal para todos.

Yücel and Rizvanoğlu [12] investigaram o impacto da empatia e das características da voz na percepção de assistentes de voz por idosos (65-75 anos) na Turquia. Utilizando uma abordagem qualitativa centrada no usuário, o estudo descobriu que um assistente de voz empático leva a uma maior abertura dos participantes, que se sentiram validados em suas experiências. Os participantes expressaram preferência por um assistente de voz que oferecesse suporte emocional e social, em vez de apenas informações.

In: IV Concurso de Trabalhos de Iniciação Científica (CTIC 2024). Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web (CTIC'2024). Juiz de Fora/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2024.  
© 2024 SBC – Sociedade Brasileira de Computação.  
ISSN 2596-1683

<sup>1</sup><https://www.notion.so/>

É possível notar a tendência de uso de LLMs em produtos voltados para idosos. A maioria das aplicações encontradas atualmente são da área da saúde, por exemplo, Ramjee et al. [9] desenvolveram um *chatbot*, treinado com uma base de dados considerada segura, que tira as dúvidas de idosos sobre a cirurgia de catarata, porém, outras aplicações têm foco em pacientes psiquiátricos idosos [6], que ajudam na comunicação entre médico e paciente [11] ou até mesmo em robôs de companhia [3].

As pesquisas demonstram um crescente interesse no desenvolvimento e aprimoramento de *chatbots* acessíveis, com foco particular na experiência do usuário. No entanto, desafios como a garantia da acessibilidade para públicos específicos, como idosos, e a aplicação ética da IA generativa, ainda precisam ser abordados de forma mais abrangente. Ao encontro dessa informação, também se destaca a ausência de material intuitivo que dê suporte ao desenvolvimento de sistemas acessíveis.

### 3 DESENVOLVIMENTO DOS ARTEFATOS TÉCNICOS

Nesta seção serão apresentados os artefatos e as metodologias empregadas no desenvolvimento dos mesmos.

#### 3.1 ARIA

O primeiro artefato técnico desenvolvido foi o *chatbot* ARIA (Assistente de Recomendação e Interação Acessível). O ARIA foi projetado para ser uma ferramenta prática que implementa as diretrizes de acessibilidade documentadas em [1], servindo como um exemplo funcional e uma plataforma de interação acessível.

O principal objetivo do ARIA é atuar como um assistente virtual voltado para a recomendação de conteúdos culturais, como filmes, séries, livros e músicas, tendo um enfoque particular na experiência dos idosos.

**3.1.1 Tecnologia Utilizada.** O ARIA foi desenvolvido utilizando a linguagem de programação Python, uma escolha popular para o desenvolvimento de *chatbots* devido à sua simplicidade e extensa variedade de bibliotecas e *frameworks*. Para a geração das respostas, o ARIA utiliza o *Gemini Pro*<sup>2</sup>, um LLM *open-source* criado pela Google, que oferece uma capacidade robusta de compreensão e geração de texto em linguagem natural.

Ao empregar o modelo de linguagem *Gemini Pro*, o ARIA é capaz de compreender contextos complexos e nuances linguísticas, além de adaptar suas respostas ao histórico de interações e preferências dos usuários, criando uma experiência personalizada e elevando a qualidade da experiência do usuário.

#### 3.1.2 Funcionalidades do ARIA.

- **Geração de Respostas Textuais:** Conforme mencionado anteriormente, o ARIA utiliza o modelo *Gemini* para gerar respostas textuais em tempo real, proporcionando interações naturais e relevantes com os usuários;
- **Transcrições em Áudio:** Além das respostas textuais, o ARIA oferece transcrições em áudio das respostas geradas. Esta funcionalidade é crucial para usuários com deficiências visuais ou dificuldades de leitura, permitindo-lhes interagir

com o *chatbot* de forma auditiva. Para realizar esta função, é utilizada a API (Interface de Programação de Aplicação ou *Application Programming Interface*, em Inglês) Text-to-Speech<sup>3</sup> da Google;

- **Possibilidade de Utilizar o Microfone:** O ARIA permite que os usuários utilizem o microfone para enviar mensagens ao *chatbot*. O áudio captado pelo microfone é transcrito em tempo real para a caixa de entrada de texto, podendo ser facilmente ajustado pelo usuário;
- **Encontrar Links de Conteúdos Culturais:** Uma das funcionalidades distintivas do *chatbot* ARIA é a sua capacidade de encontrar links de filmes, séries e músicas por meio de técnicas de *web scraping*.

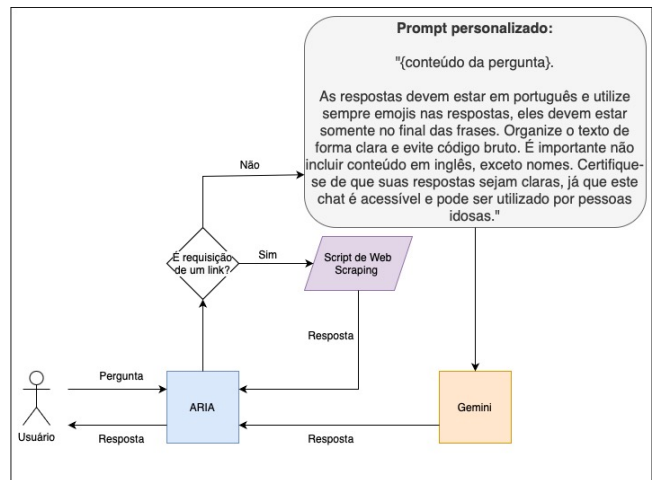


Figura 1: Fluxograma do *chatbot* ARIA.

**3.1.3 Funcionamento do ARIA.** O fluxo de funcionamento pode ser ilustrado na Figura 1. Inicialmente, o usuário envia uma mensagem na interface do ARIA. Essa mensagem pode ser escrita digitando ou utilizando o microfone que transcreve a fala do usuário.

Em seguida, a mensagem enviada para o ARIA é transmitida para o *Gemini Pro* por meio de um *prompt* personalizado, conforme ilustrado na Figura 1. Essa personalização na mensagem enviada para o LLM proporciona melhores respostas e o cumprimento das diretrizes de acessibilidade necessárias.

A resposta gerada pelo *Gemini* é retornada para o ARIA e exibida para o usuário. Além disso, também é gerado o áudio do texto retornado, o qual aparece junto da resposta textual.

No caso especial do usuário solicitar que o ARIA encontre um link de um filme, série ou música, a mensagem não é transmitida para o *Gemini*. Ao invés disso, uma função de *web scraping* é chamada, e ela retorna o link, caso seja possível encontrá-lo. Por fim, o link encontrado é exibido pelo ARIA para o usuário.

O ARIA está disponível em um repositório<sup>4</sup>, onde podem ser encontradas mais informações, bem como imagens ilustrativas do *chatbot* em funcionamento.

<sup>3</sup><https://cloud.google.com/text-to-speech>

<sup>4</sup>Para acessar o repositório, utilize o seguinte link: [https://github.com/carlosnrib/chatbot\\_v2](https://github.com/carlosnrib/chatbot_v2).

<sup>2</sup><https://deepmind.google/technologies/gemini/pro/>

### 3.2 Interactive Access

Durante a elaboração do ARIA, descrito na Seção 3.1, percebeu-se que algumas das diretrizes de acessibilidade não apresentavam o grau de clareza necessário para que fossem implementadas corretamente. Assim, a segunda etapa do desenvolvimento dos artefatos técnicos envolveu a criação de uma interface web dedicada a organizar e exibir as diretrizes. A plataforma desenvolvida tem como título *Interactive Access: Explorando Diretrizes para Chatbots Acessíveis*, e está disponível publicamente<sup>5</sup>.

**3.2.1 Tecnologia Utilizada.** O Notion foi a ferramenta escolhida para a elaboração da página, devido à sua versatilidade e capacidade de integração de diferentes tipos de conteúdo, como textos, imagens, tabelas e links, o que facilita a construção de uma plataforma abrangente e interativa.

**3.2.2 Estrutura da Página.** A página no Notion foi estruturada de forma que a apresentação das diretrizes de acessibilidade fossem feitas de duas formas de visualização: galeria principal ou pelo menu lateral. Ao clicar em uma diretriz, de ambos os modos citados, o Notion direciona o usuário para uma página específica dedicada a essa diretriz. Nessa página, são apresentadas descrições detalhadas, recomendações práticas, justificativas claras, exemplos de uso concretos e ferramentas relevantes para auxiliar na implementação.

## 4 ESTUDOS DE AVALIAÇÃO

### 4.1 Avaliação de Desempenho

Para avaliar o desempenho dos artefatos técnicos elaborados, foram conduzidos testes de carga e estresse em ambos os artefatos e o teste de tempo de resposta apenas no *chatbot*.

**4.1.1 Teste de Carga/Estresse.** Os testes de carga e estresse foram realizados em três cenários distintos para garantir a escalabilidade do sistema. Esses testes avaliaram a capacidade do *chatbot* de suportar diferentes volumes de tráfego e solicitações simultâneas. Os cenários de teste considerados estão descritos na Tabela 1.

Tabela 1: Cenário de Teste.

Cenário	Número de usuários	Taxa de crescimento	Duração
1	50	50 por seg.	5 min
2	200	10 por seg.	10 min
3	500	20 por seg.	15 min

**4.1.2 Testes de Tempo de Resposta.** Além dos testes de carga e estresse, também foram realizados testes de tempo de resposta para garantir que o *chatbot* responda de forma eficiente às solicitações dos usuários. As tarefas selecionadas para a avaliação dos tempos de resposta foram:

- Tarefa 1:** Realizar uma primeira interação com o *chatbot*.  
– **Exemplo:** "Bom dia"
- Tarefa 2:** Solicitar a recomendação de um filme de um determinado gênero.  
– **Exemplo:** "Recomende um filme de ação"
- Tarefa 3:** Solicitar uma explicação sobre um filme específico.

<sup>5</sup>O Interactive Access pode ser acessado pelo seguinte link: <https://encurtador.com.br/x4oxC>.

### – Exemplo: "Conte mais sobre o filme Toy Story"

Para executar todas as três tarefas, participaram quatro pessoas da Universidade de São Paulo, todas pesquisadoras na área de computação e do mesmo grupo de pesquisa responsável por este trabalho.

Os tempos de resposta foram calculados através de um *script* integrado na própria estrutura do *chatbot*. Em cada uma das tarefas foram consideradas três medidas distintas: tempo da geração da resposta textual, tempo da geração do áudio e tempo total.

### 4.2 Avaliação de Usabilidade

Para executar a avaliação de acessibilidade participaram três pessoas da Universidade de São Paulo, todas pesquisadoras na área de computação e do mesmo grupo de pesquisa responsável por este trabalho, sendo duas delas com experiência prévia em Interação Humano-Computador<sup>6</sup>.

Foi empregada uma avaliação de usabilidade com base nas dez heurísticas de Nielsen [7, 8], onde uma ou mais tarefas são apresentadas aos participantes que devem executá-las.

Os participantes deveriam procurar, individualmente, por problemas que firmam as heurísticas e classificá-las em um dos quatro graus de severidade:

- **Cosmético**, quando não há necessidade imediata de solução;
- **Simple**, quando o problema é considerado de baixa prioridade, isto é, pode ser reparado;
- **Grave**, nesse caso, o problema é considerado de alta prioridade e deve ser reparado;
- **Catastrófico**, quando é considerado muito grave e deve ser reparado de qualquer forma.

Após as análises individuais os participantes se reuniram usando uma plataforma de reuniões *online* para consolidar os resultados. Nessa etapa, eles deveriam conversar para discutir seus achados, suas concordâncias, discordâncias e chegarem em um acordo e finalizarem os resultados em um único documento.

## 5 RESULTADOS

Nesta seção serão apresentados os resultados obtidos por meio dos testes de carga e de usabilidade realizados com o ARIA e o Interactive Access.

### 5.1 Resultados da avaliação de desempenho

**5.1.1 Resultados do Teste de Carga/Estresse.** Foram realizados testes de carga somente no ARIA, hospedado em uma máquina virtual da Universidade de São Paulo, pois é o artefato a que temos acesso ao código. Os resultados estão dispostos na Tabela 2.

Tabela 2: Resultados do Teste de Carga/Estresse.

Cenário	Requisições	Falhas	Média (ms)	Mediana (ms)
Cenário 1	4210	1	5	5,07
Cenário 2	33720	13	4	4,64
Cenário 3	126730	59	4	4,51

<sup>6</sup>Os autores entendem a necessidade de avaliar as ferramentas com o público-alvo, entretanto, até a data da submissão deste artigo o pedido de estudo com idosos está em análise pela Plataforma Brasil.

Os resultados do teste de carga indicam que o sistema demonstra uma capacidade notável de lidar com diferentes níveis de requisições. Em todos os cenários, o sistema operou com alta confiabilidade, menos de 0,05% de falhas. Além de ser rápido e consistência, dado que as médias e medianas foram muito próximas e com valores muito pequenos. Assim, o sistema demonstrou ser capaz de suportar até 500 usuários simultâneos de forma eficiente.

**5.1.2 Resultados dos Testes de Tempo de Resposta.** Pelo mesmo motivo, foram realizados testes de tempo de resposta apenas no *chatbot*. Os resultados estão descritos na Tabela 3.

**Tabela 3: Tempo Médio por Tarefa.**

Tarefa	Mensagem (s)	Áudio (s)	Total (s)
Tarefa 1	4,93	0,84	5,89
Tarefa 2	5,85	6,17	12,14
Tarefa 3	7,32	13,25	20,69

A análise dos resultados do teste de tempo revela que o *chatbot* é capaz de responder em tempos satisfatórios para o usuário. É importante salientar que as respostas textuais são bem rápidas, e que o tempo da geração de áudio, como esperado, aumenta conforme o aumento do texto de resposta. Um ponto a se destacar é que, mesmo em uma tarefa mais complexa, como o caso da Tarefa 3, o *chatbot* responde em um tempo médio bastante aceitável, próximo de 20 segundos, proporcionando uma resposta completa e interativa, com áudio e texto, para o usuário.

## 5.2 Resultados da avaliação de usabilidade

Os resultados das avaliações de usabilidade realizadas com os artefatos serão apresentadas a seguir.

**5.2.1 ARIA.** Os avaliadores encontraram quatro problemas de usabilidade no ARIA. São eles:

- (1) Por vezes o ARIA teve algum problema de processamento que ocasionou em erro, mas não deixa isso claro ao usuário;
- (2) Em caso de erro no ARIA, independente do motivo, basta que o usuário envie alguma mensagem que o diálogo será retomado, entretanto, isso pode não ficar claro para o usuário;
- (3) Os avaliadores não encontraram estratégias que ajudassem o usuário a evitar erros;
- (4) Apesar de o usuário tirar suas dúvidas usando o próprio diálogo com o ARIA, isso não fica claro para o usuário. O ARIA também não possui documentação que auxilie na navegação.

Todos os problemas de usabilidade encontrados no ARIA foram classificados como grave e devem ser corrigidos em trabalhos futuros. Entretanto, apesar de terem sido considerados graves, não houve nenhum problema que impedisse o andamento das interações seguindo as tarefas definidas para a realização da avaliação, o que caracteriza que o ARIA possui boa usabilidade.

**5.2.2 Interactive Access.** Com a avaliação realizada, os avaliadores concordaram com a existência de dois problemas de usabilidade, apresentados a seguir.

- (1) O site não possui documentação e ajuda;
- (2) Ao clicar na diretriz usando o conteúdo central da página, é aberta uma janela menor no estilo *pop-up* e, nele, não há opção para voltar ou fechar a página.

Assim como no ARIA, os problemas de usabilidade encontrados no Interactive Access devem ser corrigidos em trabalhos futuros. Porém, os problemas relatados não interferiram na execução de nenhuma das tarefas, indicando que o Interactive Access possui boa usabilidade e, tanto o ARIA

quanto o Interactive Access, podem seguir para a etapa de avaliação com usuários idosos.

## 6 CONSIDERAÇÕES FINAIS

Este trabalho relatou o desenvolvimento e avaliação de um *chatbot* acessível que faz recomendações, ARIA, e de um site com um conjunto de diretrizes para o desenvolvimento de *chatbots* acessíveis, Interactive Access.

Os resultados encontrados foram encorajadores para a continuação da pesquisa, pois, nenhum problema grave ou que, de acordo com a experiência dos autores, influencie de forma negativa ou cause frustração nos usuários, principalmente os idosos.

A principal limitação deste trabalho é ainda não ter realizado estudos com usuários reais e em contextos de uso reais. Como trabalhos futuros, se pretende resolver os problemas de usabilidade encontrados no ARIA, realizar estudos com usuários reais e, após essa etapa, verificar se há necessidade de refinamento do *chatbot*.

## AGRADECIMENTOS

Os autores gostariam de agradecer o apoio financeiro da FAPESP (projeto número 2022/07016-9) e CNPq. O presente trabalho também foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## REFERÊNCIAS

- [1] Lucas Padilha Modesto de Araujo. 2024. *Diretrizes de apoio ao processo de construção de sistemas conversacionais de recomendação acessíveis: um estudo com usuários idosos*. Dissertação de Mestrado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos. doi:10.11606/D.55.2024.tde-16042024-104313. Recuperado em 2024-08-19, de [www.teses.usp.br](http://www.teses.usp.br).
- [2] Jiahao Chen, Fu Guo, Zenggen Ren, Xueshuang Wang, and Jaap Ham. 2023. Human-chatbot interaction studies through the lens of bibliometric analysis. *Universal Access in the Information Society* (2023), 1–20.
- [3] Iván Esteban-Lozano, Álvaro Castro-González, and Paloma Martínez. 2024. Using a LLM-Based Conversational Agent in the Social Robot Mini. In *International Conference on Human-Computer Interaction*. Springer, 15–26.
- [4] Marco Furini, Silvia Mirri, Manuela Montangero, and Catia Prandi. 2020. Do conversational interfaces kill web accessibility?. In *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1–6.
- [5] Leonid Grinin, Anton Grinin, and Andrey Korotayev. 2023. Global Aging and our Futures. *World Futures* 79, 5 (2023), 536–556.
- [6] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. [n. d.]. Evaluating and Auditing LLM-Driven Chatbots for Psychiatric Patients in Clinical Mental Health Settings. *human performance* 1, 5 ([n. d.]), 19.
- [7] Jakob Nielsen. 1995. How to conduct a heuristic evaluation. *Nielsen Norman Group* 1, 1 (1995), 8.
- [8] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256.
- [9] Pragnya Ramjee, Bhuvan Sachdeva, Satvik Golechha, Shreyas Kulkarni, Geeta Fulari, Kaushik Murali, and Mohit Jain. 2024. CataractBot: An LLM-Powered Expert-in-the-Loop Chatbot for Cataract Patients. *arXiv preprint arXiv:2402.04620* (2024).
- [10] Hasanin Mohammed Salman, Wan Fatimah Wan Ahmad, and Suziah Sulaiman. 2023. A design framework of a smartphone user interface for elderly users. *Universal Access in the Information Society* 22, 2 (2023), 489–509.
- [11] Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. 2024. Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–35.
- [12] Yeliz Yücel and Kerem Rizvanoğlu. 2022. I am old too!: Understanding the Impact of Empathy and Voice Characteristics on Older Adults' Perception of Voice Assistants. In *2022 IEEE International Conference on e-Business Engineering (ICEBE)*. IEEE, 299–304.