

“A realidade nem sempre é frustrante”: Alinhando os modelos mentais de usuários com as reais capacidades de Assistentes de Voz (AVs)

“Reality is not always frustrating”: Towards aligning users’ mental models with Voice Assistants’ (VAs) real capabilities

MOTTA, Isabela; Mestre em Design; PUC-Rio

isabela.canellas@gmail.com

QUARESMA, Manuela; Doutora em Design; PUC-Rio

mquaresma@puc-rio.br

Assistentes de Voz (AVs), como Alexa, estão cada vez mais populares, mas os modelos mentais dos usuários são incompatíveis com as reais capacidades desses sistemas. Para melhorar as interações, alinhar as percepções dos usuários com o real funcionamento dos AVs é necessário, e, para isso, é essencial considerar as percepções dos designers desses produtos. Assim, foram conduzidas entrevistas exploratórias semi-estruturadas com profissionais experientes em interfaces conversacionais, visando explorar suas opiniões sobre as causas para as lacunas nos modelos mentais dos usuários e identificar soluções para o problema. Os resultados indicaram que as características antropomórficas dos AVs e a apresentação de informações foram considerados influentes para as percepções dos usuários, assim como marketing, ficção científica e as características dos usuários. Ainda, aumentar a transparência do sistema foi considerada uma solução adequada, mas os designers precisam ter as habilidades e conhecimento para aplicar a solução, considerando as necessidades dos usuários e de usabilidade.

Palavras-chave: Assistentes de Voz; Modelos Mentais; Usabilidade.

Voice Assistants (VAs) such as Siri and Alexa are growing in popularity, but users’ mental models of VAs are not in line with the reality of these systems’ functioning. To improve interactions, aligning users’ perceptions with the VAs’ capabilities may be beneficial, and for such a work, it is essential to consider the perceptions of these products’ designers. Thus, we conducted exploratory semi-structured interviews with professionals experienced with conversational interfaces to explore their opinions on the causes for misalignments in users’ mental models and survey solutions to address this issue. The results indicate that VAs’ anthropomorphic features and information-related outputs (e.g., feedback) were considered influential for users’ perceptions, as well as marketing, science fiction, and user backgrounds. To address misperceptions, increasing VAs’ transparency was considered an adequate solution, but designers must possess the skills and knowledge to apply such a solution, considering users’ needs and usability requirements.

Keywords: Voice Assistants; Mental Models; Usability.

1 Introdução

Assistentes de Voz (AVs), como Siri e Alexa, são agentes virtuais de inteligência artificial com os quais os usuários interagem por meio de uma interface de voz, acompanhada ou não de um *display* visual (WEST; KRAUT; HAN EI, 2019). Os AVs são cada vez mais populares, e estima-se que, até 2020, essas interfaces funcionaram em mais de quatro bilhões de dispositivos (MOAR; ESCHERICH, 2020). Tais sistemas vêm sendo desenvolvidos por várias empresas de tecnologia, como Apple (Siri), Amazon (Alexa), Microsoft (Cortana), Google (Google Assistant) e Samsung (Bixby) e funcionam em diversos dispositivos, incluindo fones de ouvido, alto-falantes inteligentes e smartphones. Como demonstrado pela Alexa, que foi capaz de executar mais de 70.000 *skills* nos EUA até 2020 (STATISTA, 2020), os recursos disponíveis estão crescendo rapidamente em número, e variam de tarefas como a previsão do tempo até a automação doméstica.

Embora cada vez mais populares, algumas barreiras à adoção e ao uso de AVs ainda prevalecem. Dentre elas, estão atitudes negativas dos usuários em relação à interação de voz, preocupações com a privacidade, a percepção de baixa utilidade dos AVs e diferentes tipos de falhas no reconhecimento de comandos que levam à frustração (BURBACH et al, 2019; COWAN et al., 2017; KISELEVA et al., 2016; LOPATOVSKA et al., 2019; LUGER; SELLEN, 2016; MCLEAN; OSEI-FRIMPONG, 2019; MORIUCHI, 2019; MOTTA; QUARESMA, 2019; PRIDMORE et al., 2019; PURINGTON et al., 2017). Tais questões podem estar relacionadas aos modelos mentais dos usuários de AVs: modelos conceituais que os usuários criam para representar como um produto ou sistema funciona (NORMAN, 2013), incluindo um conjunto de expectativas sobre os componentes, funcionamento e uso adequado de um sistema (WICKENS et al., 2014). Segundo a literatura, usuários de agentes conversacionais têm modelos mentais inconsistentes (LUGER; SELLEN, 2016) e expectativas iniciais irrealistas para os AVs (CHO; LEE; LEE, 2019). Essas percepções errôneas podem levar às barreiras mencionadas anteriormente, diminuindo o potencial dos AVs para mudar os atuais paradigmas interativos de interfaces gráficas para interfaces de voz (WEST; KRAUT; HAN EI, 2019).

Para melhorar as interações usuário-AV e fomentar a adoção desses sistemas, os designers devem compreender porque existem desalinhamentos nos modelos mentais dos usuários e levantar caminhos para lidar com a questão. Assim, o objetivo deste estudo foi explorar as opiniões de especialistas em design conversacional sobre as causas para as lacunas nos modelos mentais de usuários de AVs, além de identificar soluções para lidar com o problema.

2 Modelos Mentais e Assistentes de Voz (AVs)

Modelos mentais são modelos conceituais que representam a expectativa e a compreensão dos usuários sobre o funcionamento, uso e características de um sistema (NORMAN, 2013; WICKENS et al., 2014). Esses modelos - que estão intimamente relacionados aos níveis de desempenho na realização de tarefas - se baseiam em várias fontes de informação, mas as características de design de um produto são altamente influentes para tais percepções. Wilson e Rutherford (WILSON; RUTHERFORD, 1989) explicam que os modelos mentais são formados a partir de "experiências anteriores, bem como a observação atual" (p.3, tradução nossa), reforçando que a representação de um sistema é baseada não apenas no nível de experiência de um usuário, mas também na própria interação. Da mesma forma, Norman (2013) argumenta que os usuários desenvolvem seus modelos mentais a partir da imagem do sistema: dicas físicas perceptíveis do próprio produto, experiências passadas, propagandas, manuais, etc. De acordo com Norman (2013), os designers têm seus próprios modelos

conceituais de um objeto, mas, como o produto não está mais com eles, ele se torna isolado, sendo representado apenas por sua estrutura física (ou seja, a imagem do sistema). Apesar dos projetistas desejarem que os modelos conceituais dos usuários sejam iguais aos seus, os usuários só podem fazer inferências a partir da imagem do sistema, tornando os aspectos de design vitais para suas percepções.

Embora os modelos mentais sejam essenciais ao desempenho das tarefas, a literatura mostra que os usuários têm modelos incompatíveis com os AVs (CHO; LEE; LEE, 2019; LUGER; SELLEN, 2016). Em geral, estudos indicam que os usuários têm grandes expectativas em relação à inteligência, nível de humanidade e capacidades práticas e sociais dos AVs. A literatura também aponta que essas percepções estão relacionadas com as características antropomórficas dos AVs (isto é, a capacidade de falar e seus nomes e vozes semelhantes às humanas) (CHO; LEE; LEE, 2019; KUZMINYKH et al., 2020; LUGER; SELLEN, 2016). Similarmente, vários estudos relataram que os usuários desejam - sem sucesso - realizar tarefas adicionais, como se envolver em relações emocionais e ter conversas de nível humano com os AVs (AMMARI et al., 2019; LOPATOVSKA et al., 2020).

Além disso, as evidências sugerem que os usuários têm baixa compreensão do funcionamento dos AVs, incluindo informações sobre o funcionamento do sistema, formas de uso e políticas de privacidade. A literatura mostra que os usuários subutilizam ou abandonam tarefas ao longo do tempo, relatam confusão sobre fontes de erros e ações do sistema, têm preocupações com a privacidade e empregam métodos de aprendizado ineficientes (BENETEAU et al., 2020; CHO; LEE; LEE, 2019; LUGER; SELLEN, 2016; PRIDMORE et al., 2019). A causa dessas percepções errôneas pode estar na forma como os AVs apresentam as informações. Foi demonstrado que os usuários utilizam as respostas dos AVs (ex: feedback, instruções) para raciocinar e interpretar a interação, identificar erros e reformular comandos (PORCHERON et al., 2018; PORCHERON; FISCHER; SHARPLES, 2017). Em linha com essa possibilidade, a literatura mostra que maiores níveis de explorabilidade (ou *discoverability*; ex: fornecer instruções) podem melhorar a usabilidade de agentes conversacionais (KIRSCHTHALER; PORCHERON; FISCHER, 2020), e respostas mais transparentes podem levar a uma melhor compreensão das fontes de erro (KIM; JEONG; LEE, 2019; MYERS et al., 2018).

Considerando a importância das características de um produto (isto é, a imagem do sistema) para os modelos dos usuários, os AVs devem ser projetados de forma que suas características representem adequadamente suas funcionalidades e destaquem suas formas de uso. Portanto, é primordial identificar as causas que levam às lacunas nos modelos mentais de usuários de AVs e encontrar soluções de design para mitigar esse problema. Tal exploração pode ser beneficiada ao compreender as percepções dos designers de interfaces conversacionais, uma vez que os aspectos de design do produto derivam das conceptualizações do próprio designer (NORMAN, 2013). Portanto, embora a literatura tenha trazido conhecimento essencial sobre as perspectivas dos usuários, investigar as opiniões dos designers sobre questões que levam a lacunas nos modelos dos usuários e como resolvê-las pode destacar percepções relevantes para potencializar as interações usuário-AV.

3 Metodologia

Considerando o exposto acima, realizamos entrevistas exploratórias com profissionais experientes na pesquisa e/ou desenvolvimento de interfaces conversacionais. As entrevistas tiveram como objetivo responder duas perguntas de pesquisa (PP):

PP1: Quais as causas que esses profissionais atribuem aos modelos mentais desalinhados dos usuários de AVs?

PP2: Que soluções estes profissionais acreditam ser eficazes e adequadas para resolver a questão dos modelos mentais desalinhados de usuários de AVs?

3.1 Participantes

Os participantes foram recrutados por e-mail e mídia social (LinkedIn), seguindo o método de amostragem de bola de neve e amostragem por conveniência. Recrutamos tanto pesquisadores da academia, quanto desenvolvedores de agentes conversacionais. Para participar, os pesquisadores deveriam ter o título de doutor, enquanto os desenvolvedores deveriam ter pelo menos três anos de experiência em projetos de interfaces conversacionais. Todos os participantes deveriam estar atualmente ou recentemente envolvidos em estudos ou projetos sobre interfaces conversacionais. Para diversificar ao máximo as visões dos entrevistados, recrutamos participantes de diferentes formações (tabela 1). No total, oito profissionais participaram das entrevistas.

Tabela 1 – Características dos participantes

Característica	Participantes
Gênero	Masculino = 5 Feminino = 3
País	Brasil = 7 Reino Unido = 1
Local de trabalho	Mercado = 6 Academia = 1 Ambos = 1
Formação	Ciência da computação = 2 Comunicação = 2 Design = 1 Biblioteconomia = 1 Produção multimídia = 1 Letras = 1

Fonte: as autoras.

3.2 Procedimento

Primeiramente, agendamos as entrevistas e enviamos um termo de consentimento livre e esclarecido aos participantes por e-mail. As entrevistas começaram com uma apresentação superficial do objetivo do estudo e, então, seguiu o roteiro abaixo:

- **Explicar o conceito de modelos mentais.** Para alinhar a compreensão do conceito por todos os participantes, a entrevistadora explicou a definição de modelos mentais utilizada no estudo (baseado em Norman, 2013 e Wickens et al., 2014): "um tipo de modelo conceitual criado pelos usuários para representar como um produto ou

sistema funciona, incluindo uma série de expectativas sobre seus componentes, funcionamento e uso adequado".

- **Explicar a questão dos modelos mentais de usuários de AVs.** A entrevistadora explicou que a literatura sugere que os modelos mentais dos usuários de AVs não estão alinhados com as capacidades reais desses sistemas, resultando em abandono e subutilização.
- **Perguntar sobre as causas de modelos mentais desalinhados e soluções para lidar com tal questão.** A entrevistadora pediu aos participantes que expressassem suas opiniões sobre o que leva os usuários a formar modelos mentais desalinhados com a realidade. Os participantes também deveriam dar sugestões sobre como resolver esse problema. A entrevistadora encorajou os participantes a darem exemplos de situações reais com base em sua experiência de trabalho, sempre que possível.
- **Explicar e pedir aos participantes que comentem sobre a suposição de que as características antropomórficas e os outputs dos AVs influenciam os modelos mentais dos usuários.** Como apresentado na seção anterior, a literatura sugeriu duas causas principais para lacunas nos modelos dos usuários: antropomorfismo e respostas do sistema (ou *outputs*). Escolhemos incluir esses tópicos na entrevista para explorar as opiniões dos participantes sobre se, e como, tais características dos AVs influenciam os modelos dos usuários. Primeiramente, a entrevistadora explicou sua suposição sobre esses dois temas e, então, pediu aos profissionais que dessem sua opinião sobre tal pressuposto. Este tópico foi colocado após a pergunta anterior propositalmente para que os especialistas pudessem dar suas opiniões imparciais antes de comentar o assunto.
- **Explicar e pedir aos participantes que comentem sobre a suposição de que aumentar a transparência dos AVs melhoraria os modelos mentais dos usuários¹.** Similarmente ao tópico anterior, a literatura indica que aumentar os níveis de transparência dos AVs (ou seja, fornecer mais instruções e explicações) poderia diminuir a lacuna nos modelos mentais dos usuários. Por isso, a entrevistadora apresentou a suposição de que aumentar a transparência do sistema, mantendo níveis semelhantes de humanidade nos AVs, alinharia os modelos mentais dos usuários com as capacidades reais do sistema. Os participantes deviam comentar se acreditam que tal solução seria viável e desejável.
- **Agradecimentos.**

As entrevistas foram realizadas remotamente por meio da plataforma Zoom. O procedimento durou em média 45 minutos e foi gravado em áudio e vídeo para análise. As entrevistas foram realizadas entre junho e julho de 2021.

3.3 Análise

Para analisar os dados, realizamos uma análise temática seguindo uma abordagem *top-down*. Em primeiro lugar, após a transcrição das entrevistas, identificamos como as respostas dos profissionais se enquadram nas seguintes categorias:

- Causas para os modelos mentais desalinhados com as reais capacidades dos AVs.
- Soluções para lidar com as lacunas nos modelos mentais de usuários de AVs.

¹ Para uma versão expandida dos resultados e discussão sobre o tópico do aumento da transparência em respostas de AVs, veja a publicação prévia das autoras: MOTTA; QUARESMA, 2022.

Em seguida, identificamos tópicos emergentes para cada categoria, transpondo trechos para a plataforma Miro e elaborando um diagrama de afinidade (BARNUM, 2011).

4 Resultados

4.1 Causa de problemas nos modelos mentais de usuários de AVs

Os especialistas indicaram quais fatores eles acreditavam impactar os modelos mentais de usuários de AVs e gerar percepções desalinhadas sobre esses sistemas.

Em primeiro lugar, conforme evidenciado pela literatura, os participantes consideraram que os AVs possuem um alto nível de humanidade, afetando a forma como os usuários percebem o produto. Alguns participantes mencionaram essa causa antes de pedirmos diretamente suas opiniões, enquanto outros não a mencionaram inicialmente. No entanto, todos os participantes concordaram que as características antropomórficas dos AVs influenciam os modelos mentais dos usuários. Eles mencionaram que tal antropomorfismo leva os usuários a acreditar que podem interagir com os AVs de maneira semelhante às conversas humanohumano. *“Eu acho que direcionar as pessoas para esse entendimento, ‘olha, você pode falar como se estivesse conversando com um amigo’, isso é um erro”* (P3). De acordo com os profissionais, a percepção dos usuários se transforma em uma percepção mais realista gradualmente, à medida que erros e falhas acontecem.

Os resultados das entrevistas também estão em consonância com estudos anteriores sobre os aspectos que induzem tal humanidade. Os especialistas explicaram que características de design, como voz humana (ou seja, voz gravada em vez de conversão de texto para fala), estilo conversacional (por exemplo, menos robótico e mais natural) e vocabulário podem gerar alta humanidade nos AVs. Um participante mencionou que uma voz menos robótica pode induzir expectativas mais baixas em relação às capacidades do sistema: *“Quando usamos um TTS [Text-To-Speech], as pessoas acabam entendendo (...) Fica mais claro que, tudo bem, é uma tecnologia, aqui é uma máquina, então pode ser um pouco mais limitado”* (P6).

De forma inesperada, os especialistas argumentaram que atribuir níveis tão altos de humanidade a interfaces conversacionais pode até fazer os usuários questionarem se o sistema é realmente uma máquina ou um humano. Alguns participantes relataram exemplos de usuários que confundiram um agente conversacional com um humano, fazendo com que se sentissem enganados e constrangidos: *“Algumas pessoas sabem que estão interagindo com uma máquina, então, às vezes, até se sentem um pouco enganados [quando a interface tenta imitar um humano]. (...) Mas acontece, algumas pessoas interagem e demoram um pouco mais para perceber que é uma máquina, e quando descobrem ficam super envergonhadas”* (P2). Essa confusão também parece afetar o desempenho de tarefas: *“Às vezes, algumas pessoas menos atentas não entendiam desde o início que era uma gravação [e não um humano]. Às vezes, eles contavam toda a história para o agente digital, e era inútil porque ele não conseguia entender. Só precisávamos que eles nos dissessem se podiam pagar [uma dívida] ou não”* (P6). Para um dos especialistas, aumentar a humanidade dos AVs a ponto de os usuários não terem mais certeza se é uma máquina ou um humano também é uma questão ética.

Em linha semelhante ao tópico anterior, os especialistas comentaram como esse antropomorfismo pode fazer os usuários esperarem conversas em nível humano, confundindo-os sobre como interagir. Os participantes argumentaram que, como a comunicação humana diária leva a uma noção bem estabelecida de como conversar, os usuários acreditam que devem imitar essas práticas de conversação ao interagir com

interfaces altamente humanas. *“Se estou falando com um ser humano, [sei] como interagir com um ser humano, aprendemos isso através da experiência durante nossas vidas. Se a máquina se apresenta para mim como algo o mais próximo possível de um humano, entendo que, para interagir com ela, devo interagir mais ou menos da mesma forma que interajo com humanos”* (P3). No entanto, quando a interface não atende a essas expectativas devido às suas limitações, os usuários não conseguem interagir e se sentem frustrados. *“A ideia que vem à nossa mente é ‘vou ter uma conversa’. Mas, às vezes, até os próprios humanos não respondem da maneira que você imaginou.”* (P4).

Embora os efeitos adversos da humanidade dos AVs na percepção dos usuários tenham sido relatados na literatura, as entrevistas indicaram que a aplicação de tais aspectos antropomórficos não é impulsionada apenas por decisões de design. Os especialistas mencionaram que parceiros de negócios, gerentes ou clientes intervínham em projetos de interfaces conversacionais frequentemente. Esses atores acreditavam que um agente virtual mais humano tornaria as interações mais naturais e melhoraria tanto a experiência, quanto a aceitação do usuário. *“Eles queriam que humanizássemos as aplicações, e ‘humanizar’ vem disso, de entender que quanto mais próximo o sistema atuasse como humano, mais fluida [a interação] seria. E melhores seriam as reações [dos usuários] e sua aceitação do sistema.”* (P3). No entanto, dois especialistas também mencionaram que esses stakeholders geralmente não possuem o conhecimento técnico para aplicar as soluções de forma adequada: *“Pode ser o cliente, o empresário, às vezes é um executivo que nem sabe nada [sobre o projeto e AVs] mas quer dar pitaco. (...) [Pessoas que] não têm a menor ideia do que estão falando, mas dão suas opiniões.”* (P5).

Outro tema emergente relacionado aos achados da literatura são sobre os *outputs* dos AVs. Embora nem todos os especialistas tenham mencionado aspectos relacionados aos *outputs* dos AVs em um primeiro momento, observamos uma concordância geral com a relevância desse fator. Os especialistas argumentaram que a falta de informações causa problemas na percepção dos usuários, uma vez que as interfaces conversacionais, frequentemente, não apresentam o que o sistema é capaz de fazer e nem como utilizá-lo. Um participante argumentou que os AVs não fazem uma preparação inicial para os usuários: *“Se você inicia a interação com esses sistemas sem preparar mais ou menos os usuários para tais interações, seja com uma explicação inicial, seja com uma forma mais sutil de mostrar que eles estão conversando com um sistema que tem formas de interação específicas, você já está distorcendo a percepção deles de que [o sistema] tem essas possibilidades”*. (P3). Como a interface não esclarece o que ela pode fazer, os usuários acabam explorando a interface por conta própria, utilizando uma abordagem de tentativa e erro. Um participante também argumentou que os mecanismos de recuperação de erros dos AVs geralmente são limitados e não fornecem informações relevantes aos usuários, que não podem fazer perguntas para investigar as fontes de erros: *“[Quando você fala com um humano] você pode fazer esse tipo de interrogatório. Mas quando você fala com uma interface conversacional, ela diz: ‘desculpe, eu não entendi’, e você não consegue[perguntar] ‘O que você não entendeu?’”* (P8).

Os especialistas também mencionaram a influência das características dos usuários em suas percepções sobre os AVs. Tais características incluíam tanto suas experiências anteriores com outras interfaces de voz, quanto suas formações acadêmicas ou interesses pessoais. Um participante mencionou que alguns usuários se sentem tensos desde o início de uma interação por voz devido a experiências ruins anteriores com interfaces de voz. Com base em sua pesquisa, outro especialista argumentou que os usuários se comportam diferentemente

dependendo de seu nível de experiência e interesse em tecnologia. “[As pessoas interessadas em tecnologia] entendem o funcionamento da tecnologia, o ambiente e as possibilidades. (...) Às vezes, eles até [interagem] para testar [o AV]. (...) [Pessoas menos interessadas em tecnologia] usam apenas os recursos mais básicos. Eles podem até obedecer o que a interface está tentando ensinar, repetir exatamente as mesmas palavras” (P6). Um profissional argumentou que a questão da percepção equivocada dos usuários sobre os AVs é natural, pois esses sistemas são uma tecnologia relativamente nova, e sempre há um estágio de estranheza: “Acho natural porque definitivamente não é novo, mas ainda está no começo, reconhecendo como lidar com toda essa expectativa” (P7).

Por fim, os seguintes tópicos emergentes são causas que – assim como as características dos usuários – não estão inseridas no próprio AV, mas pertencem ao seu contexto de uso: marketing dos AVs e a percepção dos usuários sobre as máquinas em geral. Segundo os especialistas, as propagandas dos AVs são irrealistas, provocando percepções incorretas nos usuários, principalmente, em relação à capacidade de conversação dos AVs. Em seus pontos de vista, as empresas atribuem termos como “humano” ou “conversacional” ao AV para criar a imagem de uma interação fluida, semelhante à comunicação humana. “Há muito tempo, cerca de oito anos, a Apple dizia: ‘você pode falar com a Siri como falaria com seu amigo, e funciona’. Claro, isso não é verdade agora e, certamente, não era verdade naquela época.” (P8). Um participante também argumentou que o próprio termo “inteligência artificial” pode causar equívocos: “Acho que o termo inteligência artificial complica muito porque a palavra ‘inteligência’ traz o imaginário de que [o AV] é, de fato, inteligente” (P5). Da mesma forma, alguns profissionais acreditavam que as percepções de maior inteligência e capacidade podem ser devidas não apenas ao marketing dos AVs, mas também à impressão geral das pessoas de que as máquinas são perfeitas e altamente capazes. “Às vezes atribuímos um nível de perfeição às máquinas, e não fazemos isso quando falamos com outras pessoas” (P4).

4.2 Soluções para melhorar os modelos mentais de usuários de AVs

Esta seção apresenta as opiniões dos participantes sobre quais soluções podem ser implementadas para lidar com problemas nos modelos mentais de AVs dos usuários.

Em primeiro lugar, enquanto os especialistas consideraram que as características antropomórficas dos AVs são um motivador relevante para as percepções incompatíveis dos usuários, suas opiniões sobre como lidar com esse aspecto não foram homogêneas. Por um lado, os profissionais reconheceram o valor de um agente humanizado para tornar as interações mais naturais. Um participante comentou que humanizar a forma de falar dos AVs também pode beneficiar a apresentação das informações para os usuários, fazendo a ponte entre usuários e máquinas: “Quando se fala em ‘explicabilidade’, acho que o antropomorfismo ajuda um pouco. Você pode usar uma linguagem que talvez seja mais acessível, que não seja a linguagem de uma máquina.” (P2). Da mesma forma, um especialista apontou que esses aspectos antropomórficos são importantes para a imagem dos AVs: “Obviamente, você não vai vender um assistente virtual mal-humorado, chato e com uma voz estranha que ninguém quer ouvir”. (P4). Além disso, um participante comentou que retirar a humanização dos AVs seria um desafio devido à forte associação entre seres humanos e conversas: “Retirar a humanização é quase impossível porque, mesmo que a voz seja robótica, já estamos trazendo um elemento humano. Não há outros seres que falam. (...) Talvez a gente consiga diminuir [a humanização], mas não tirar” (P5).

Por outro lado, os participantes ponderaram como implementar os níveis corretos de humanidade nos AVs, argumentando que exagerar algumas características pode ser mais

prejudicial do que benéfico. Muitos especialistas mencionaram que os desenvolvedores devem privilegiar a objetividade e a eficiência das interações em detrimento das características antropomórficas, como um estilo de conversa lúdico e dicas engraçadas (ex: piadas). Em linha semelhante, um participante sugeriu que os desenvolvedores deveriam escolher os momentos certos para apresentar respostas lúdicas aos usuários: *“Você precisa saber o momento certo para ser humano na interface (...) talvez no início ou no final (...), não durante a interação.”* (P5). Comentários nesse sentido destacaram que os objetivos dos usuários na interação devem ser investigados e priorizados: *“O que você espera? Se você espera [realizar uma tarefa], realmente vale a pena eu fazer [o AV] tão humanizado que confunde a pessoa? Talvez seja melhor se for um pouco mais objetivo, certo?”* (P2).

O segundo tópico emergente estava relacionado a como os *outputs* dos AVs são projetados. Segundo os participantes, um AV deve apresentar mais informações sobre seu funcionamento aos usuários, incluindo esclarecimentos explícitos ou implícitos de que o agente conversacional não é um ser humano. Dois especialistas relataram casos em que empregaram dicas instruindo os usuários sobre como interagir para sugerir que o agente era uma máquina. Tais instruções também foram mencionadas por outros profissionais, que argumentaram que as interfaces conversacionais devem explicar o que o sistema pode fazer ou como funciona para alinhar o entendimento dos usuários. Os profissionais sugeriram que interfaces mais objetivas e instrutivas levam a um melhor desempenho das tarefas: *“Agora que estamos trabalhando com microcrédito, a primeira resposta do chatbot é apresentar seu escopo (...) Quanto mais direcional, mais proativo, melhor funciona”* (P2). No entanto, um participante argumentou que decidir quanta informação é suficiente para os usuários entenderem o sistema é uma tarefa desafiadora para os desenvolvedores. *“Quanto as pessoas precisam saber para formar seus modelos mentais? (...) Interfaces de voz são relativamente novas, então não sabemos como explicá-las particularmente bem e quais explicações precisamos dar às pessoas. (...) Talvez não precisemos entrar em detalhes, ou talvez precisemos.”* (P8).

Da mesma forma, os participantes também argumentaram que projetar estratégias adequadas de recuperação de erros é fundamental para ajudar os usuários a se recuperarem de eventuais erros. *“Quando um erro acontece, você precisa mostrar ao usuário o motivo de por que aconteceu. Isso ajuda a pessoa a responder a pergunta pela segunda vez, a reformular sua resposta sem não repetir o mesmo erro.”* (P6). No entanto, um especialista estava preocupado que ensinar diretamente aos usuários exatamente o que falar poderia prejudicar a interação.

Além de lidar com os níveis de humanidade e apresentar informações aos usuários, quando perguntamos aos especialistas quais soluções eles aplicaram em trabalhos anteriores, alguns mencionaram que é preciso atentar para as diretrizes gerais do design e as boas práticas do design conversacional. *“Existem algumas regras básicas com princípios de design que devem ser aplicadas em qualquer tipo de interação”* (P3). Essas recomendações orientaram as decisões desde a estrutura da conversa, como o tipo de pergunta a ser aplicada (por exemplo, escolha aberta ou múltipla), até um design mais detalhado em nível de conteúdo.

Da mesma forma, os profissionais recomendaram que considerar o contexto dos usuários e os tipos de sistema é crucial para a implementação de soluções. Os desenvolvedores devem considerar quem são os usuários e em quais situações, para quais finalidades e como usarão o AV. *“Precisamos considerar o contexto em que isso será abordado. O público, os usuários, suas personas, a tecnologia.”* (P1). Alguns dos especialistas também argumentaram que, devido a essa dependência do contexto, seria difícil fornecer soluções gerais e absolutas para a questão do modelo mental. De forma consistente, muitos especialistas relataram que, embora tentem

prever uma possível ação do usuário que possa levar a falhas, para corrigir erros e aplicar soluções, geralmente precisam examinar a interface e lidar com cada problema separadamente. *“Acho que há essa primeira análise de qual é o erro (...), e depois propomos uma solução.”* (P6).

Ainda no âmbito das soluções, os participantes deram suas opiniões sobre se aumentar a transparência nas respostas dos AVs seria uma solução viável e desejável. Enquanto alguns especialistas mencionaram que a solução parecia tecnicamente viável, outros se mostraram céticos sobre o assunto. Um profissional explicou que os AVs exigem múltiplos sistemas e tecnologias para funcionar, levando a várias fontes potenciais de falhas. Por isso, o sistema nem sempre é capaz de diagnosticar por que um erro ocorreu, e nem explicar como revertê-lo. Os profissionais relataram, ainda, que podem haver limitações projetuais para o aumento da transparência, uma vez que o projeto de interfaces conversacionais está muito ligado à construção de fluxos conversacionais. Segundo esses participantes, prever cenários e incluir *outputs* de alta transparência em fluxos previamente estabelecidos pode ser um desafio. Os entrevistados também expressaram preocupação com as limitações decorrentes da habilidade da equipe de desenvolvimento e eventuais demandas dos parceiros de negócios, como mencionado na seção 4.1.

Também perguntamos aos participantes se eles acreditavam que aumentar a transparência nos *outputs* dos AVs era desejável, e a maioria considerou a solução vantajosa, principalmente pela usabilidade dos AVs. De modo geral, aqueles que julgaram a proposta como desejável comentaram que alavancar a transparência poderia resultar em menos erros, tornar os usuários mais confiantes e seguros, e melhorar o entendimento sobre os limites da máquina: *“Faz sentido ser honesto com o usuário (...) As pessoas entendem que, tudo bem, é uma máquina, não vai fazer tudo (...) Claro que tem a questão da expectativa e da realidade, mas a realidade nem sempre é frustrante”* (P4). Ainda, os especialistas mencionaram que aumentar a transparência dos AVs pode ajudar não apenas os usuários, mas também os demais envolvidos no próprio projeto do sistema. Porém, outros profissionais expressaram preocupação de que um aumento excessivo na transparência pudesse impactar negativamente a usabilidade dos AVs, tornando as interações lentas, tediosas, confusas ou difíceis.

5 Discussão

Esta seção irá abordar as perguntas de pesquisa à luz da literatura: **PP1:** Quais causas esses profissionais atribuem aos modelos mentais desalinhados de AVs dos usuários? **PP2:** Quais soluções esses profissionais acreditam ser eficazes e adequadas para resolver a questão dos modelos mentais desalinhados de AVs dos usuários?

Em primeiro lugar, os especialistas mencionaram que os AVs são altamente humanizados, afetando a percepção dos usuários a um nível que pode confundi-los sobre como interagir e impactar o desempenho das tarefas. Esses resultados estão de acordo com a literatura que mostra como características antropomórficas, como nomes e vozes humanizadas, podem gerar expectativas irreais nos usuários (LUGER; SELLEN, 2016). No entanto, remover completamente a humanidade dos AVs pode não ser possível tanto do ponto de vista do design, quanto de negócio. Nass e Moon (2000) argumentaram que mesmo aspectos humanos sutis inerentes às interfaces de voz (ex: voz, fala) podem provocar antropomorfismo e, consequentemente, um comportamento de conversação inconsciente nos usuários. Além disso, estudos apontam que os traços de personalidade dos AVs são motivo de adoção de AVs

(LUGER; SELLEN, 2016; MOTTA; QUARESMA, 2019) e são particularmente valorizados por nichos de usuários, como crianças (FESTERLING; SIRAJ, 2020). Similarmente, os especialistas comentaram que *stakeholders* fora da equipe de desenvolvimento solicitam agentes conversacionais humanizados. Assim, as características antropomórficas dos AVs trazem benefícios e não podem ser descartadas, mas devem ser usadas com cuidado para não atrapalhar as tarefas principais.

Em segundo lugar, os especialistas mencionaram que os AVs apresentam poucas informações sobre seu funcionamento e não oferecem bom suporte para a recuperação de erros, gerando confusão sobre a operação dos AVs. Como solução, os profissionais expressaram que os AVs devem apresentar informações sobre o que podem fazer e como se recuperar de erros. No geral, aumentar a transparência foi considerado uma solução desejável para mitigar erros e aumentar a confiança dos usuários. Esses achados corroboram evidências que mostram que os usuários utilizam as respostas dos AVs para raciocinar e interpretar as ações do sistema (PORCHERON et al., 2018; PORCHERON; FISCHER; SHARPLES, 2017) e que sistemas mais transparentes estão relacionados a um melhor desempenho e influenciam a compreensão sobre o erro fontes (KIM; JEONG; LEE, 2019; KIRSCHTHALER; PORCHERON; FISCHER, 2020; MYERS et al., 2018). Independentemente das vantagens e desvantagens da solução, os profissionais entrevistados demonstraram ceticismo quanto à viabilidade de aumentar a transparência dos AVs. Por exemplo, limitações tecnológicas na tecnologia de IA que suportam AVs podem impedir que a interface diagnosticique e explique corretamente as falhas. Sobre esse tema, a literatura sobre *eXplainable Artificial Intelligence* (Inteligência Artificial Explicativa; XAI) tem indicado que muitos sistemas de IA – principalmente aqueles com melhor desempenho – são caixas pretas, incapazes de explicar como tomam decisões (HOLZINGER, 2018). Os especialistas também relataram que projetar transparência em interfaces conversacionais exigiria a compreensão prévia do fluxo conversacional, mas como AVs são interfaces generalistas - isto é, realizam tarefas em diversas situações e domínios -, prever contextos de uso é desafiador.

Apesar das vantagens de interações mais explícitas, expor quantidades excessivas de informações pode tornar as interações chatas e lentas, prejudicando sua objetividade e usabilidade. Para equilibrar cuidadosamente usabilidade e transparência e empregar outras soluções de design, os especialistas consideraram vital que os desenvolvedores se preocupassem com boas práticas de interação por voz e tivessem as habilidades apropriadas para esse trabalho. Dentre tais conhecimentos, atentar-se ao contexto dos usuários (ex: características, antecedentes, localização, etc.) foi considerado primordial. Por exemplo, fatores externos ao design do AV, como marketing e as influências culturais da mídia de ficção científica, foram considerados impactantes para os modelos mentais dos usuários. Da mesma forma, os profissionais mencionaram idade, experiência em tecnologia e uso prévio de AVs como influentes para a percepção dos usuários. Conforme demonstrado na literatura, usuários com alta experiência com AVs e tecnologia são mais bem-sucedidos nas interações e entendem melhor o funcionamento dessas interfaces (LAU; ZIMMERMAN; SCHAUB, 2018; LOPATOVSKA et al., 2019; LUGER; SELLEN, 2016; MYERS; FURQAN; ZHU, 2019). Assim, como sugerido por alguns profissionais, as quantidades de informações a serem exibidas poderiam ser definidas – entre outros parâmetros – pelo nível de experiência dos usuários com AVs e tecnologia.

Os AVs estão cada vez mais populares e trazem diversos benefícios aos usuários. No entanto, a adoção dos AVs ainda é dificultada por barreiras relacionadas aos modelos mentais dos usuários, que não estão alinhados com as reais capacidades desses sistemas. Para melhorar as interações, os designers devem entender as causas desse problema e empregar soluções adequadas para preencher a lacuna entre os modelos dos usuários e a realidade dos AVs. O objetivo deste estudo foi explorar as opiniões de especialistas em design conversacional sobre as causas dos desalinhamentos nos modelos mentais de usuários de AVs e identificar soluções para resolver esse problema. Para isso, realizamos entrevistas exploratórias com profissionais experientes com pesquisa ou desenvolvimento de interfaces conversacionais.

Os achados trouxeram indicações sobre as causas gerais que levam a lacunas nos modelos mentais de usuários de AVs. Enquanto alguns fatores estavam relacionados ao design dos AVs (por exemplo, humanidade, respostas do sistema), outros fatores externos também causavam percepções errôneas (por exemplo, marketing, histórico do usuário). Quanto a como lidar com a questão, as soluções podem diferir dependendo dos domínios de tarefas disponíveis no AV e seus contextos de uso. Considerar o contexto do usuário é essencial ao escolher como projetar soluções que não prejudiquem as necessidades dos usuários e os requisitos de usabilidade. Portanto, pode ser necessário investigar o contexto e as soluções de design, aplicando uma abordagem caso a caso. Esse trabalho é essencial para preencher a lacuna entre os modelos mentais dos usuários e as reais capacidades desses sistemas, mitigando expectativas irrealas e confusões que podem levar à frustração. Atualmente, os AVs têm uma ampla gama de tarefas e recursos disponíveis e, como mencionado por um especialista, a realidade do que os AVs têm a oferecer não precisa ser frustrante.

Este estudo limitou-se a uma amostra de apenas oito profissionais, e estudos posteriores podem se beneficiar de uma amostra maior e mais variada. No entanto, as entrevistas seguiram uma abordagem exploratória, sendo parte inicial de uma pesquisa maior que visa investigar possíveis problemas que levam a desalinhamentos nos modelos mentais dos usuários e levantar soluções para lidar com o assunto.

Agradecimentos

O presente trabalho foi realizado com apoio da FAPERJ e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

7 Bibliografia

- AMMARI, T.; KAYE, J.; TSAI, J. Y.; BENTLEY, F. Music, Search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction*, vol. 26, no. 3, 2019. DOI 10.1145/3311956.
- BARNUM, C. **Usability Testing Essentials. Ready, Set...** Burlington, MA: Elsevier, 2011. DOI 10.1016/B978-0-12-375092-1.00023-4.
- BENETEAU, E.; GUAN, Y.; RICHARDS, O. K.; ZHANG, M. R.; KIENTZ, J. A.; YIP, J.; HINIKER, A. Assumptions Checked: How Families Learn About and Use the Echo Dot. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, New York, NY, USA, vol. 4, no. 1, Mar. 2020. DOI 10.1145/3380993.
- BURBACH, L.; HALBACH, P.; PLETTENBERG, N.; NAKAYAMA, J.; ZIEFLE, M.; CALERO VALDEZ, A. "Hey, Siri", "Ok, Google", "Alexa". Acceptance-Relevant Factors of Virtual Voice-Assistants.

2019-July., Jul. 2019. **2019 IEEE International Professional Communication Conference (ProComm) [...].** [S. l.]: IEEE, Jul. 2019. vol. 2019-July, p. 101–111. DOI 10.1109/ProComm.2019.00025.

CHO, M.; LEE, S.; LEE, K.-P. Once a Kind Friend is Now a Thing. 18 Jun. 2019. **Proceedings of the 2019 on Designing Interactive Systems Conference [...].** New York, NY, USA: ACM, 18 Jun. 2019. p. 1557–1569. DOI 10.1145/3322276.3322332.

COWAN, B. R.; PANTIDI, N.; COYLE, D.; MORRISSEY, K.; CLARKE, P.; AL-SHEHRI, S.; EARLEY, D.; BANDEIRA, N. “What can i help you with?” 4 Sep. 2017. **Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services [...].** New York, NY, USA: ACM, 4 Sep. 2017. p. 1–12. DOI 10.1145/3098279.3098539.

FESTERLING, J.; SIRAJ, I. Alexa, What Are You? Exploring Primary School Children’s Ontological Perceptions of Digital Voice Assistants in Open Interactions. **Human Development**, vol. 64, no. 1, p. 26–43, 2020. DOI 10.1159/000508499.

HOLZINGER, A. From Machine Learning to Explainable AI. Aug. 2018. **2018 World Symposium on Digital Intelligence for Systems and Machines (DISA) [...].** [S. l.]: IEEE, Aug. 2018. p. 55–66. DOI 10.1109/DISA.2018.8490530.

KIM, J.; JEONG, M.; LEE, S. C. “Why did this voice agent not understand me?” 21 Sep. 2019., e1; i1. **Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings [...].** New York, NY, USA: ACM, 21 Sep. 2019. p. 146–150. DOI 10.1145/3349263.3351513.

KIRSCHTHALER, P.; PORCHERON, M.; FISCHER, J. E. What Can i Say?: Effects of Discoverability in VUIs on Task Performance and User Experience. 2020., cited By 2. **ACM International Conference Proceeding Series [...].** [S. l.: s. n.], 2020. DOI 10.1145/3405755.3406119.

KISELEVA, J.; WILLIAMS, K.; JIANG, J.; HASSAN AWADALLAH, A.; CROOK, A. C.; ZITOUNI, I.; ANASTASAKOS, T. Understanding User Satisfaction with Intelligent Assistants. 2016. **Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval [...].** New York, NY, USA: Association for Computing Machinery, 2016. p. 121–130. DOI 10.1145/2854946.2854961.

KUZMINYKH, A.; SUN, J.; GOVINDARAJU, N.; AVERY, J.; LANK, E. Genie in the Bottle. 21 Apr. 2020. **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems [...].** New York, NY, USA: ACM, 21 Apr. 2020. p. 1–13. DOI 10.1145/3313831.3376665.

LAU, J.; ZIMMERMAN, B.; SCHaub, F. Alexa, Are You Listening? **Proceedings of the ACM on Human-Computer Interaction**, New York, NY, USA, vol. 2, no. CSCW, p. 1–31, Nov. 2018. DOI 10.1145/3274371.

LOPATOVSKA, I.; GRIFFIN, A. L.; GALLAGHER, K.; BALLINGALL, C.; ROCK, C.; VELAZQUEZ, M. User recommendations for intelligent personal assistants. **Journal of Librarianship and Information Science**, vol. 52, no. 2, p. 577–591, 8 Jun. 2020. DOI 10.1177/0961000619841107.

LOPATOVSKA, I.; RINK, K.; KNIGHT, I.; RAINES, K.; COSENZA, K.; WILLIAMS, H.; SORSCHE, P.; HIRSCH, D.; LI, Q.; MARTINEZ, A. Talk to me: Exploring user interactions with the Amazon Alexa. **Journal of Librarianship and Information Science**, vol. 51, no. 4, p. 984–997, 7 Dec. 2019. DOI 10.1177/0961000618759414.

LUGER, E.; SELLEN, A. “Like Having a Really Bad PA”: The Gulf between User Expectation and Experience of Conversational Agents. 2016. **Proceedings of the 2016 CHI Conference on**

Human Factors in Computing Systems [...]. New York, NY, USA: Association for Computing Machinery, 2016. p. 5286–5297. DOI 10.1145/2858036.2858288.

MCLEAN, G.; OSEI-FRIMPONG, K. Hey Alexa ... examine the variables influencing the use of artificial intelligent in-home voice assistants. **Computers in Human Behavior**, vol. 99, p. 28–37, 2019. DOI <https://doi.org/10.1016/j.chb.2019.05.009>.

MOAR, J.; ESCHERICH, M. Hey Siri, how will you make money? 2020. Available at: <https://www.juniperresearch.com/whitepapers/hey-siri-how-will-you-make-money>. Accessed on April 18th 2022.

MORIUCHI, E. Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. **Psychology & Marketing**, vol. 36, no. 5, p. 489–501, 15 May 2019. DOI 10.1002/mar.21192.

MOTTA, I.; QUARESMA, M. OPPORTUNITIES AND ISSUES IN THE ADOPTION OF VOICE ASSISTANTS BY BRAZILIAN SMARTPHONE USERS. **Ergodesign & HCI**, vol. 7, no. Especial, p. 138, 31 Dec. 2019. DOI 10.22570/ergodesignhci.v7iEspecial.1312.

MOTTA, I; QUARESMA, M; "Explorando desafios e oportunidades para aumentar a transparência de Assistentes de Voz (AVs)", p. 983-994 . In: **Anais do 18º Congresso Internacional de Ergonomia e Usabilidade de Interfaces Humano-Tecnologia e o 18º Congresso Internacional de Ergonomia e Usabilidade de Interfaces e Interação Humano-Computador**. São Paulo: Blucher, 2022.

MYERS, C.; FURQAN, A.; NEBOLSKY, J.; CARO, K.; ZHU, J. Patterns for how users overcome obstacles in Voice User Interfaces. 2018-April., 2018. **Conference on Human Factors in Computing Systems - Proceedings [...].** [S. l.: s. n.], 2018. vol. 2018-April, p. 1–7. <https://doi.org/10.1145/3173574.3173580>.

MYERS, C. M.; FURQAN, A.; ZHU, J. The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface. 2 May 2019., E. **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems [...].** New York, NY, USA: ACM, 2 May 2019. p. 1–9. DOI 10.1145/3290605.3300277.

NASS, C.; MOON, Y. Machines and Mindlessness: Social Responses to Computers. **Journal of Social Issues**, vol. 56, no. 1, p. 81–103, 2000. .

NORMAN, D. **The design of everyday things**. Revised an. New York, NY, USA: Basic books, 2013.

PORCHERON, M.; FISCHER, J. E.; REEVES, S.; SHARPLES, S. Voice Interfaces in Everyday Life. 2018-April., 21 Apr. 2018., i2. **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems [...].** New York, NY, USA: ACM, 21 Apr. 2018. vol. 2018-April, p. 1–12. DOI 10.1145/3173574.3174214.

PORCHERON, M.; FISCHER, J. E.; SHARPLES, S. "Do Animals Have Accents?" 25 Feb. 2017., E. **Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing [...].** New York, NY, USA: ACM, 25 Feb. 2017. p. 207–219. DOI 10.1145/2998181.2998298.

PRIDMORE, J.; ZIMMER, M.; VITAK, J.; MOLS, A.; TROTTIER, D.; KUMAR, P. C.; LIAO, Y. Intelligent Personal Assistants and the Intercultural Negotiations of Dataveillance in Platformed Households. **Surveillance & Society**, vol. 17, no. 1/2, p. 125–131, 31 Mar. 2019. DOI 10.24908/ss.v17i1/2.12936.

PURINGTON, A.; TAFT, J. G.; SANNON, S.; BAZAROVA, N. N.; TAYLOR, S. H. "Alexa is my new BFF": Social roles, user satisfaction, and personification of the Amazon Echo. **Conference on Human Factors in Computing Systems - Proceedings**, vol. Part F1276, p. 2853–2859, 2017. <https://doi.org/10.1145/3027063.3053246>.

STATISTA. Total number of Amazon Alexa skills in selected countries as of January 2020. 2020. Available at: <https://www.statista.com/statistics/917900/selected-countries-amazon-alexa-skill-count/>. Accessed on: 2 Nov. 2021.

WEST, M.; KRAUT, R.; HAN EI, C. **I'd blush if I could**, 2019. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>. Accessed on: 2 Nov 2021.

WICKENS, C. D.; LEE, J.; LIU, Y.; BECKER, S. G. Cognition. **An Introduction to Human Factors Engineering**. second. London: Pearson, 2014. p. 100–134.

WILSON, J. R.; RUTHERFORD, A. Mental models: Theory and application in human factors. **Human Factors**, vol. 31, no. 6, p. 617–634, 1989. <https://doi.org/10.1177/001872088903100601>.