

Sistemas conversacionais: aspetos culturais influenciam a avaliação de UX de um protótipo conceptual de assistente de voz

Juliana Camargo¹, José Nunes¹, Maria João Antunes¹, Óscar Mealha¹,
Carolina Abrantes² e Luís Nóbrega²

**julianacamargo@ua.pt; jnunes@ua.pt; mariajoao@ua.pt; oem@ua.pt;
cabrantes@broadvoice.com; lnobrega@broadvoice.com**

¹ DigiMedia, Universidade de Aveiro, Aveiro, Portugal

² GoTelecom, S.A, Aveiro, Portugal

DOI: 10.17013/risti.50.118–135

Resumo: Este artigo analisa a experiência de 100 indivíduos (50 brasileiros e 50 portugueses) no contexto de um sistema conversacional com assistente de voz automatizado. A avaliação da UX foi realizada em diferentes etapas: cálculo da estatística descritiva, análise das frases ditas pelos utilizadores e aplicação da escala SUS. Esta abordagem metodológica permitiu verificar que o protótipo conceptual simula um produto com grande potencial para facilitar o dia a dia dos profissionais que atuam em salões de estética e beleza. No entanto, há funcionalidades que devem ser redesenhadas, como a escolha dos nomes dos botões e a acessibilidade da ferramenta. Outra contribuição do estudo é a confirmação de que há opiniões divergentes entre as populações da amostra, brasileira e portuguesa, mostrando que a solução teve melhor aceitação entre os portugueses. Esta última questão é um dos aspetos mais relevantes dos achados deste trabalho, pois pode comprometer o sucesso de tal assistente.

Palavras-chave: sistemas conversacionais, chatbots, assistentes virtuais, inteligência artificial, acessibilidade, experiência do utilizador (UX).

Conversational systems: cultural aspects influence the UX evaluation of a conceptual voice assistant prototype

Abstract: This paper analyzes the experience of 100 individuals (50 Brazilians and 50 Portuguese) in the context of a conversational system with an automated voice assistant. The UX evaluation consisted of calculating descriptive statistics, analyzing the phrases spoken by users during the tests and applying the SUS scale. This methodological approach made it possible to verify that the conceptual prototype simulates a product with great potential for facilitating the daily lives of professionals working in beauty salons. However, there are features that need to be redesigned, such as the choice of button names and the accessibility of the tool. Another contribution is the confirmation that there are divergent opinions between

the Brazilian and Portuguese sample populations, showing that the solution was better accepted by the Portuguese. This last issue is one of the most relevant aspects of this study's findings, as it could jeopardize the success of such an assistant.

Keywords: conversational systems, chatbots, virtual assistants, artificial intelligence, accessibility, User Experience (UX).

1. Introdução

A avaliação da Experiência do Utilizador (UX) tem ganhado cada vez mais relevância na investigação académica (Lallemand & Koenig, 2020). Tal disciplina consiste em verificar o comportamento de um grupo de pessoas durante a utilização inicial de um produto, sistema ou serviço (Pettersson et al., 2018). É importante ressaltar que esse tipo de estudo vai além da observação das respostas dadas pelos participantes durante os testes: com uma metodologia adequada, é possível detetar características hedônicas e emocionais, como estética, estímulo e identificação, que são extremamente relevantes em qualquer tipo de interação entre humanos e sistemas (Bernhaupt & Pirker, 2014).

Uma avaliação bem estruturada, portanto, não se limita à verificação da usabilidade do produto ou da solução a ser testada. Ela deve ser construída a partir de uma combinação de técnicas capazes de analisar todos os aspetos envolvidos na jornada do utilizador (Ludwig et al., 2021). Trata-se de uma atividade multidisciplinar, uma vez que envolve várias áreas do conhecimento, como design, engenharia e psicologia (Pettersson et al., 2018).

Quando nos referimos a um sistema conversacional, tema central deste estudo, a análise de UX é importante para validar recursos e propor melhorias (Guerino & Valentim, 2020). Este tipo de sistema consiste no uso de tecnologias capazes de compreender a linguagem humana e conduzir uma conversa entre pessoas e dispositivos, de forma natural, que pode acontecer por meio da escrita ou da fala (Trippas et al., 2020). Alguns exemplos mais recorrentes são os assistentes pessoais, como Siri e Alexa, e os chatbots, muito utilizados no atendimento ao cliente.

A utilização mais ampla desta tecnologia tem sido amplamente suportada por avanços significativos nas áreas de Processamento de Linguagem Natural (PLN), Inteligência Artificial (IA) e Aprendizado de Máquina. Estas abordagens permitem uma interação mais fluida e consistente entre seres humanos e computadores, justificando seu uso em situações cotidianas (Kiseleva et al., 2016). Considerando que estas abordagens são sustentadas pelas necessidades e preferências dos utilizadores, o software depende de uma avaliação de UX consistente para identificar se a solução proposta é adequada. Isto geralmente é feito por meio de técnicas específicas, como entrevistas, observação, narrativas e diários.

O objetivo deste artigo é, portanto, relatar e discutir os resultados da avaliação de UX de um protótipo conceptual de um sistema de conversação. Este estudo empírico contou com uma amostra de 100 participantes, sendo metade da população brasileira e metade composta por cidadãos portugueses. O paper foi organizado da seguinte forma: o primeiro passo foi a revisão da literatura. Em seguida, na segunda secção, há uma breve descrição do protótipo conceptual e suas diferentes iterações de design. A terceira secção

relata o teste de experiência do utilizador, ou seja, quais técnicas foram empregadas para a coleta de dados e sua análise. Por fim, a quarta secção apresenta os resultados e sua discussão, com destaque para os principais achados e contribuições deste trabalho.

2. Estudos relacionados

Antes de entrar na análise dos métodos utilizados para avaliar os sistemas de conversação, é importante esclarecer a diferença entre Experiência do Utilizador (UX) e usabilidade. Esta última, segundo a norma ISO 9241: Parte 11, mede a eficácia, a eficiência e a satisfação de um produto num contexto de utilização específico, com base num conjunto pré-determinado de objetivos e tarefas (International Organization for Standardization, 2010). Relativamente à experiência do utilizador, de acordo com a ISO 9241-110:2010 (cláusula 2.15), trata-se da análise das respostas geradas a partir da utilização inicial de um produto, sistema ou serviço. Por conseguinte, são conceitos distintos, precisamente porque os dados da avaliação de UX dependem essencialmente de emoções, sentimentos e perceções de cada participante durante a utilização de um produto.

A recolha de informações sobre a experiência do utilizador pode ocorrer de várias formas: por meio de entrevistas, observações, questionários validados ou uma combinação de técnicas. Um exemplo de abordagem metodológica mista é o estudo realizado por Abreu et al. (2021) para avaliar a UX de um sistema de interação por voz no contexto da televisão interativa (iTV). Os investigadores optaram por uma triangulação de instrumentos: AttrakDiff, System Usability Scale (SUS) e Self-Assessment Manikin (SAM), complementados por entrevistas semiestruturadas.

Iniesto et al. (2021) também apostaram em uma combinação de técnicas para avaliar a UX de um assistente virtual destinado a estudantes com deficiência física. Os investigadores utilizaram a observação direta, questionários (aplicados antes e depois das atividades) e entrevistas semiestruturadas. Além disso, pediram aos participantes que preenchessem dois questionários quantitativos: SUS e SUISQ-R, uma versão simplificada do questionário Speech User Interface Service Quality (SUISQ).

Outro exemplo foi conduzido por Brüggemeier et al. (2020), que adotaram escalas validadas para avaliar a UX referente à Alexa, a assistente pessoal da Amazon, especificamente em termos de seleção de músicas. Este estudo utilizou os questionários Subjective Assessment of Speech System Interfaces (SASSI), SUISQ-R, SUS e AttrakDiff. Ao longo do processo de avaliação, os participantes foram questionados sobre quais instrumentos consideravam mais eficazes para medir ou classificar as suas perceções, tendo o questionário SASSI sido identificado como a escolha preferida. O estudo começou com um teste inicial para identificar os dados pessoais e o estado emocional dos participantes. Posteriormente, utilizaram o Computer System Usability Questionnaire (CSUQ), que se baseia na usabilidade.

Procedimentos semelhantes foram realizados por Ludwig et al. (2021) e Le Pailleur et al. (2020), que objetivaram identificar aspetos emocionais relacionados a um sistema inteligente empregando uma combinação de instrumentos, como o questionário SAM e entrevistas semiestruturadas.

Ainda sobre a melhoria da experiência do utilizador, o estudo conduzido por Watearachchi et al. (2022) teve como objetivo desenvolver um modelo de dispositivo móvel que possa tomar decisões acertadas – e em tempo real – sobre funcionalidades com base nas emoções e no contexto do utilizador. Os investigadores aplicaram diversos inquéritos e, a partir das respostas dos utilizadores, criaram protótipos utilizando a técnica de prototipagem evolutiva. O modelo foi considerado aceitável como um sistema adaptativo com controlo do utilizador, capaz de sugerir funções com base nas emoções e no contexto.

Importa ressaltar que a análise destes trabalhos aqui apresentados foi crucial para apoiar as escolhas dos instrumentos utilizados na avaliação da UX relativa ao protótipo desenvolvido no âmbito deste estudo, cuja descrição pode ser encontrada na secção seguinte.

3. Protótipo

O protótipo conceptual apresentado neste artigo é derivado do projeto CallBob - Voice Bots Builder, que funciona como um serviço de mediação de chamadas telefónicas. O objetivo foi o desenvolvimento de um software adaptado a pequenas e médias empresas, permitindo-lhes receber e gerir marcações (por voz) automaticamente, via chamadas telefónicas. A partir deste objetivo, foi criado um protótipo de assistente automatizado utilizando o software InVision¹.

O produto, portanto, simula a capacidade de responder a ligações automaticamente e realizar tarefas específicas, como agendar horários e fornecer informações sobre os serviços oferecidos. A gestão de conteúdo é realizada pelos próprios empreendedores em uma plataforma que pode ser acedida via internet. Cada gerente é responsável por administrar os elementos de fala do robô e definir o conjunto de respostas a serem dadas aos clientes durante suas interações com o sistema.

O protótipo foi projetado a partir de um processo de investigação participativa, incorporando opiniões de profissionais de várias áreas. Estas informações foram coletadas por meio de grupos focais e entrevistas individuais com clientes e gerentes de pequenas empresas. Com base nos dados empíricos obtidos, foram desenvolvidas narrativas de comunicação que funcionaram como suporte para que os diálogos entre humanos e o sistema fossem eficazes e naturais. O resultado é um modelo de interação intuitivo para a configuração do assistente virtual, voltado a utilizadores com níveis de literacia digital de baixa a média.

4. Protocolo e amostra

A amostra do estudo é composta por proprietários e/ou profissionais que atuam em salões de beleza. Para simular a utilização da ferramenta, a equipa de investigação propôs seis cenários com base nas solicitações feitas com maior frequência pelos clientes (identificadas pelos empreendedores que participaram dos grupos focais iniciais). Tais

¹ Disponível em: <https://www.invisionapp.com/>.

cenários e tarefas estavam relacionados com as principais funcionalidades do software, incluindo: 1) gestão do fluxo inicial de serviços, 2) configuração da área de agendamento, 3) personalização dos serviços oferecidos, 4) gestão de pedidos, 5) definição do horário de funcionamento e localização e 6) configuração das informações dos clientes. O método utilizado para analisar a experiência do utilizador em relação aos cenários envolveu uma combinação de técnicas capazes de identificar múltiplos aspetos, como estimulação, satisfação, controlo e aceitação:

- Questionário inicial: os dados pessoais e as informações profissionais foram recolhidos por meio de um questionário;
- Observação direta: os participantes utilizaram o protótipo enquanto narravam os seus pensamentos e forneciam as suas perceções;
- Avaliação dos cenários: no final da apresentação de cada cenário, os participantes classificaram a sua experiência numa escala nominal de 1 a 6, em que 1 indicava uma experiência “altamente negativa” e 6 indicava uma experiência “altamente positiva”;
- Escala de usabilidade do sistema (SUS): todos os participantes preencheram um questionário SUS, que mede a usabilidade do protótipo;
- Avaliação qualitativa do discurso dos participantes: os comentários, sugestões, elogios e observações relevantes feitos pelos participantes foram categorizados com base nos seguintes critérios: “altamente negativo” (-2), “negativo” (-1), “neutro” (0), “positivo” (1) e “altamente positivo” (2).

Os testes foram realizados online, por meio de um software de videoconferência, de fevereiro a maio de 2020, devido ao contexto pandémico. Um total de 100 profissionais foram convidados a participar do estudo, sendo 50 de Portugal e 50 do Brasil. A escolha da amostra se deu por conveniência. Dos 100 participantes selecionados, 63% são mulheres e 37% homens, com uma média de idade de 37 anos. Todos os indivíduos partilhavam a característica comum de trabalharem em salões de beleza.

5. Avaliação do protótipo e descobertas

Os resultados desta secção apresentam as classificações de experiência dos participantes (numa escala nominal de 6 pontos) para cada um dos 6 cenários que lhes foram apresentados. As estatísticas descritivas das classificações de experiência para cada cenário estão representadas na **Figura 1** como um diagrama de radar. O diagrama de radar inclui os valores do primeiro quartil, do terceiro quartil e da mediana da amostra (n=100) para cada um dos 6 cenários que constituem o protótipo conceptual.

O diagrama de radar na **Figura 1** mostra que as perceções gerais dos participantes estavam, na sua maioria, alinhadas à uma experiência “altamente positiva”. A classificação mediana para todos os cenários apresentados foi de 5. No entanto, os participantes expressaram uma opinião mais negativa especificamente em relação ao cenário 6, que diz respeito à secção de informações sobre o cliente. Este aspeto apresentou a menor amplitude entre o primeiro quartil (Q1), com 4 (limite exterior do branco interior), e o terceiro quartil (Q3), com 5 (limite exterior do sector azul). Em geral, os participantes reconheceram a utilidade desta funcionalidade para os salões.

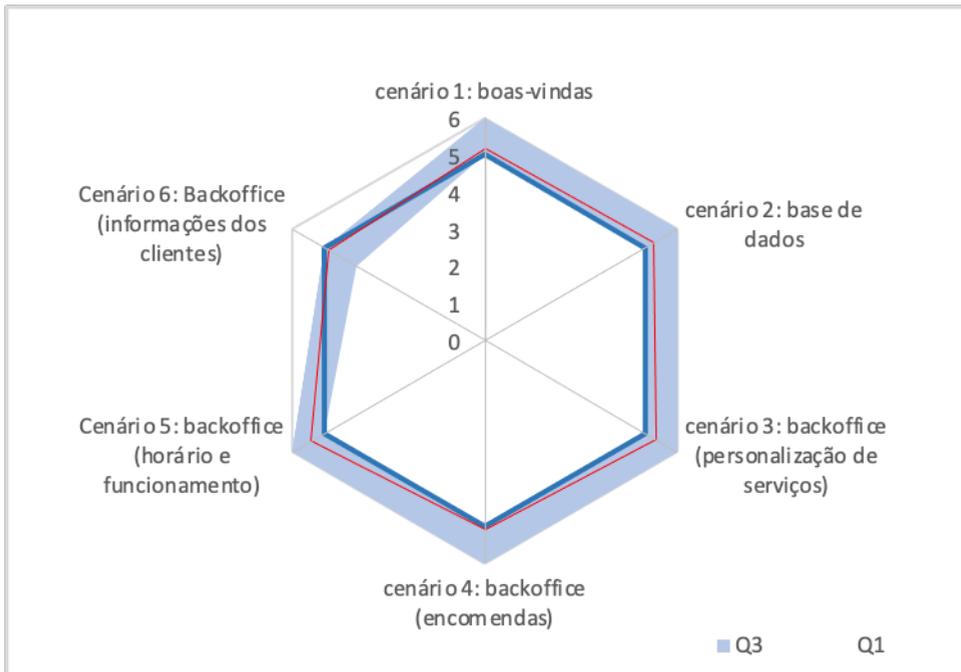


Figura 1 – Mediana, primeiro quartil (Q1) e terceiro quartil (Q3) das pontuações atribuídas pelos participantes a cada cenário proposto.

No entanto, 44 participantes sugeriram que o recurso poderia recolher dados mais pormenorizados sobre os clientes. Os seus comentários incluíam afirmações como: “Faltam informações relevantes, como o tipo de corte de cabelo feito e os produtos utilizados” e “Gosto muito da área de informações, mas acho que poderia ser mais detalhada, com a data de nascimento, por exemplo, e o tipo de corte de cabelo”. Destas 44 sugestões, 41 foram feitas por participantes brasileiros, indicando uma disparidade de opiniões entre os participantes das duas nacionalidades. Isto pode ser visto na **Figura 2** e na **Figura 3**, com os participantes brasileiros a exibirem uma reação menos positiva do que os participantes portugueses em todos os cenários, especialmente no 6. Uma possível razão para esta diferença é que 46 participantes brasileiros mencionaram que as marcações por telefone já não são o método mais comum utilizado nos salões do país. Afirmaram ainda que dependem atualmente de marcações feitas pela aplicação WhatsApp, o que pode ter contribuído para uma certa resistência em relação a algumas das funcionalidades apresentadas. Um participante comentou: “Seria mais interessante ter uma integração deste sistema com o WhatsApp, permitindo-nos enviar notificações pela aplicação.” Este contraste não foi observado entre os participantes portugueses, que indicaram utilizar prioritariamente o telefone para marcação de consultas.

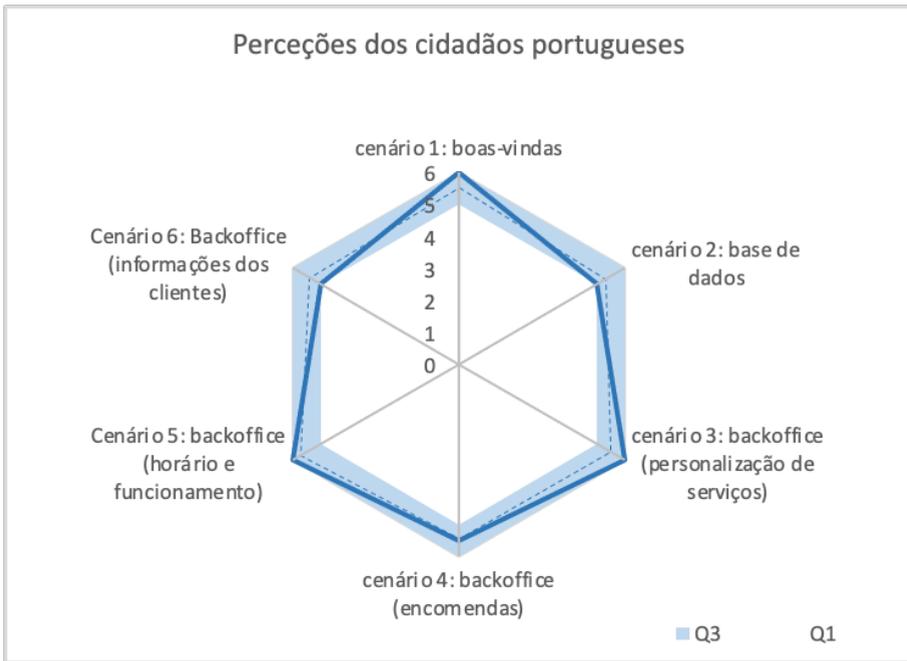


Figura 2 – Perceção dos cidadãos portugueses sobre os cenários apresentados.

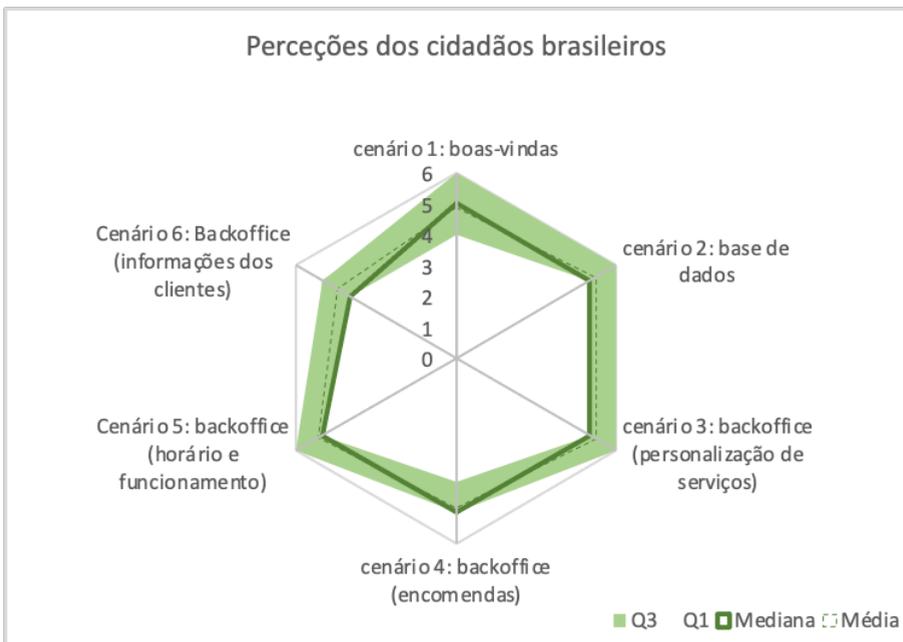


Figura 3 – Perceções dos cidadãos brasileiros sobre os cenários apresentados.

Relativamente ao questionário SUS, que foi utilizado para avaliar a usabilidade do software, a pontuação obtida foi de 80,83, indicando uma classificação entre “bom” e “excelente”. Isso significa que os participantes consideraram a ferramenta fácil de usar, mas também identificaram áreas de melhoria. Esta conclusão é apoiada pela análise individual de cada teste, em que os participantes forneceram críticas e sugestões para o sistema.

A secção seguinte apresenta os resultados destas análises, com base na recolha e organização das opiniões dos participantes. Foram analisadas 1.128 frases, classificadas nas seguintes categorias “altamente negativas” (em casos de fortes críticas ou dificuldades na realização das tarefas), “negativas” (críticas moderadas e pequenas dificuldades na realização das tarefas), “neutras” (aspectos não positivos ou negativos mencionados), “positivas” (tarefas realizadas sem ajuda) e “altamente positivas” (tarefas facilmente realizadas com elogios).

5.1. Gestão do fluxo de serviço inicial (cenário 1)

A tarefa de gestão do fluxo inicial do serviço consistia em selecionar as mensagens que seriam comunicadas pelo bot de voz aos clientes. Durante os testes, dois aspetos foram mencionados com frequência pelos participantes: “UX Writing” (mencionada 100 vezes) e “Design gráfico” (mencionada 62 vezes). A maioria (84%) deu um feedback positivo relativamente à “UX Writing”. Isto indica que boa parte dos entrevistados considerou que as mensagens apresentadas nesta etapa eram claras e concisas, melhorando a interação com o sistema (**Figura 4**).

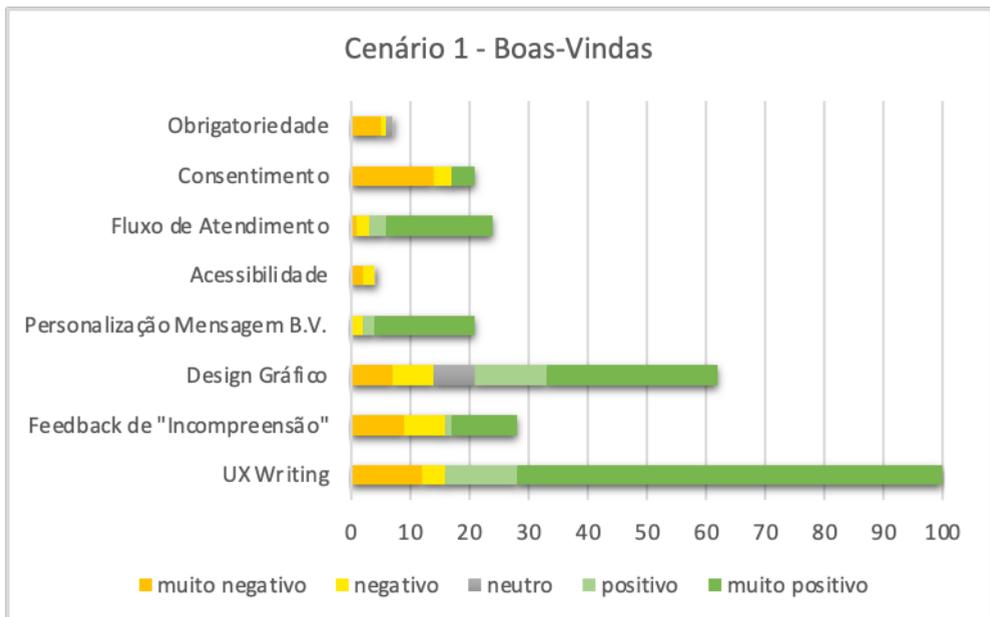


Figura 4 – Percepções sobre o cenário 1, que consistia em configurar o fluxo de boas-vindas.

Do total de 100 menções, apenas 15% expressaram aspetos negativos relativamente à gestão do fluxo inicial do serviço. Por exemplo, os participantes referiram ter dúvidas ou dificuldades em compreender certas partes do processo, como a inclusão de uma mensagem de boas-vindas ou a forma de começar a utilizar o sistema. Foram feitas sugestões no sentido de fornecer instruções mais claras ou um tutorial para facilitar a utilização.

Relativamente ao aspeto “design gráfico”, 46,7% (29) das menções foram “muito positivas” e 20% (7) foram “positivas”. Os participantes apreciaram a representação visual do fluxo do serviço, afirmando que esta proporcionava uma compreensão clara do funcionamento do sistema e das informações a preencher. No entanto, 11% (7) das menções foram classificadas como “negativas” e 11% (7) como “muito negativas”, principalmente devido a uma perceção de falta de atratividade do design. Adicionalmente, 11% (7) das menções foram consideradas neutras.

No cenário 1, três aspetos se destacaram com maior incidência de feedback negativo: “feedback de incompreensão”, “acessibilidade” e “consentimento”. Em relação ao “feedback de incompreensão”, que se refere à mensagem que o bot emitiria quando não entendesse a entrada do cliente, houve 28 menções. Destas, 57,1% (16) foram negativas e 42,8% (12) foram positivas. Os participantes que deram feedback negativo expressaram confusão quanto ao número de mensagens a registar no sistema, uma vez que havia espaço para três.

Todas as menções (4) sobre a acessibilidade da plataforma foram negativas, sendo 50% (2) classificadas como “negativas” e 50% (2) como “muito negativas”. De um modo geral, os participantes consideraram o tamanho do texto demasiado pequeno, o que compromete a legibilidade.

Relativamente ao aspeto “consentimento”, que consistia em criar uma mensagem para notificar os clientes sobre a utilização dos seus dados, houve 21 menções. Apenas 19% (4) dessas menções foram classificadas como “positivas”. Os participantes manifestaram dificuldades em compreender o conceito de consentimento e em identificar a mensagem de consentimento, o que levou a sugestões de explicações mais claras.

5.2. Configuração da área de marcação de consultas (cenário 2)

Nesta área, os utilizadores têm de registar as informações obrigatórias que o bot deve pedir aos clientes antes de marcar uma consulta. Para além disto, existe um campo dedicado à inclusão de frases para formação do sistema. A categoria “pedidos de informação dos clientes” foi a mais mencionada pelos participantes, com um total de 90 menções (**Figura 5**). Dos que comentaram, 95% (86) acharam fácil incluir as informações obrigatórias, como nome, data, serviço proposto e horário. Os 5% restantes (4) tiveram uma dúvida comum sobre a diferença entre as opções “obrigar” e “ativar” disponíveis na interface.

O campo destinado à inclusão de frases para auxiliar no treinamento do sistema foi bem recebido pela maioria dos participantes. Das 39 menções, 74% (29) foram classificadas como “positivas”. Os participantes reconheceram a importância da inclusão de novas

frases, especialmente nos casos em que interagem com clientes que falam de forma diferente devido a variações regionais. Por exemplo, um participante mencionou ter muitos clientes brasileiros em Portugal e apreciou a capacidade de acomodar diferentes padrões de fala. No entanto, 26% (10) das menções foram negativas, principalmente devido à falta de compreensão da funcionalidade. Tal aspeto influenciou outro, classificado como “terminologia do botão”. Este tema foi mencionado 38 vezes, sendo que apenas 1 opinião foi positiva, indicando que a palavra “treino” não foi a melhor escolha para o botão ou que deveria haver alguma explicação adicional junto ao ícone. Do total de menções negativas, 73% (27) foram feitas por entrevistados brasileiros, mostrando que a variação linguística entre Brasil e Portugal pode impactar a experiência do utilizador.

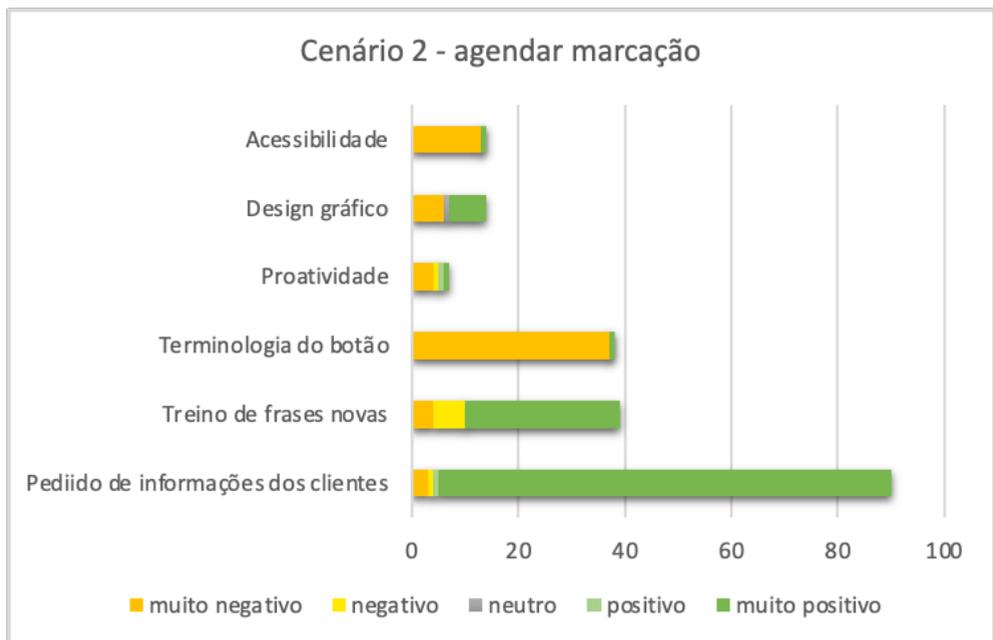


Figura 5 – Percepções em relação ao cenário 2, que consistia em configurar a área de marcações.

Outro aspeto mencionado 7 vezes foi a “proatividade”. Os participantes que o abordaram sentiram falta de uma componente proativa na plataforma. O “design gráfico”, por sua vez, foi citado 14 vezes, sendo 7 vezes “muito positivo”, 1 “neutro” e 6 “muito negativo”, mostrando que há espaço para melhorias. Alguns exemplos de sugestões dadas pelos participantes: “deveria ter uma cor diferente para cada opção do fluxo” e “o fluxo de atendimento deveria ser numerado, para que pudéssemos seguir os passos”.

Por fim, a categoria “acessibilidade” foi mencionada 14 vezes, sendo que em 92% (13) das incidências o tamanho da fonte de letra foi questionado. “Deveria ser maior para facilitar a leitura”, disse um dos participantes.

5.3. Personalização dos serviços oferecidos (cenário 3)

Esta tarefa envolveu a adição de um novo serviço oferecido ao sistema. Os participantes consideraram este procedimento fácil de executar, tendo apenas uma pessoa tido dificuldades em encontrar a localização correta (**Figura 6**).

Outro aspeto que foi frequentemente mencionado nos testes foi a relevância da funcionalidade. No total, foram 37 menções, todas elas avaliadas como “muito positivas”. O aspeto da “terminologia dos botões” apareceu 14 vezes, com 7 menções “muito positivas” e 7 menções “muito negativas”. Os participantes consideraram que a escolha de “informações comerciais” como rótulo do botão não representava corretamente a funcionalidade.

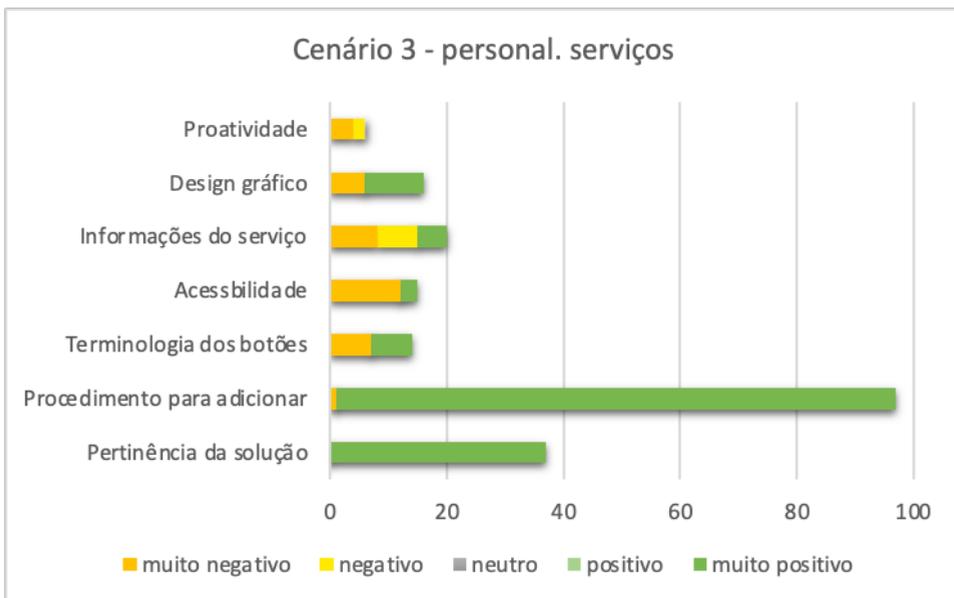


Figura 6 – Opiniões sobre o cenário 3, que consistia em incluir um novo serviço no sistema.

Assim como nos demais cenários, o aspeto “acessibilidade” foi mencionado pelos participantes. Das 15 citações, 80% (12) são “muito negativas”, referindo-se novamente ao tamanho escolhido para a fonte. “Deveria ter uma opção para aumentar as letras quando fosse mais conveniente”, disse um dos participantes.

No que diz respeito ao preenchimento das informações sobre o serviço a ser incluído, 15 entrevistados disseram que poderia haver mais campos para detalhá-lo. Por isso, consideraram este aspeto como “negativo” (7 pessoas) e “muito negativo” (8 pessoas).

5.4. Gestão das encomendas (cenário 4)

Neste cenário, os participantes foram encarregados de gerir uma encomenda efetuada por um cliente. O procedimento de marcação foi bem-sucedido, como indicado por

73 participantes que forneceram comentários classificados como “muito positivos”. Relativamente à “pertinência da solução”, apesar de ter sido elogiada por 25% (25/100) dos participantes, 12% (12/100) manifestaram preocupações quanto ao facto de poder não funcionar eficazmente (**Figura 7**). Apontaram, por exemplo, o risco de os clientes não se lembrarem do nome correto do produto quando fazem uma encomenda, uma vez que o sistema não sugere opções de forma proativa. Essa opinião foi mais prevalente entre os participantes brasileiros, representando 82% (9/12) das avaliações.

Entre os 32% (32/100) que questionaram o botão que leva à lista de pedidos, 81% (26/32) criticaram a escolha da terminologia. Entre estes, 88% (22/26) eram brasileiros, destacando mais uma vez os potenciais problemas relacionados com a variação linguística. Apenas 4% dos entrevistados sugeriram a inclusão de uma nova secção com o histórico de encomendas dos clientes.

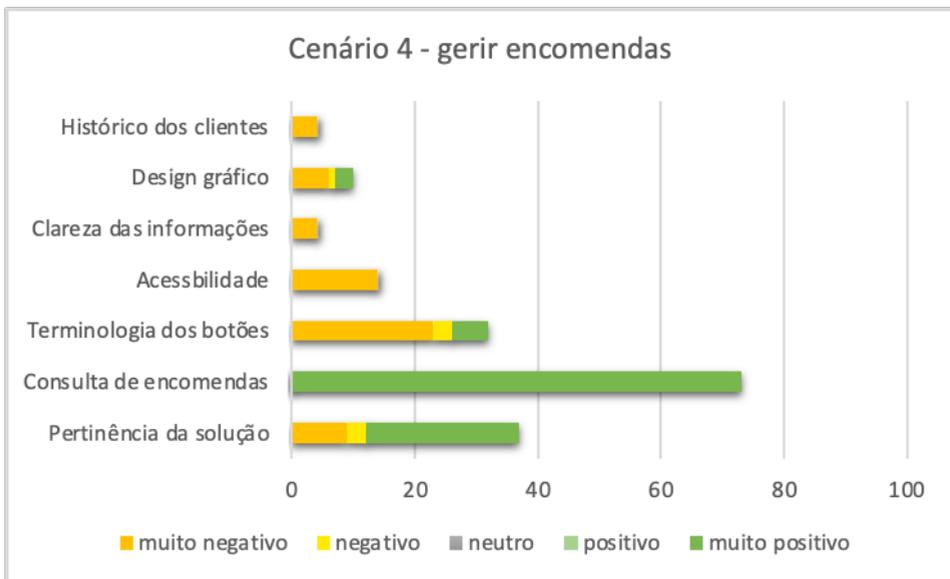


Figura 7 – Percepções sobre o cenário 4, que consistia na consulta de encomendas feitas pelos clientes.

5.5. Informações sobre horários e dias de funcionamento (cenário 5)

A tarefa atribuída para este cenário consistia em adicionar dados importantes ao sistema, como o horário de funcionamento e a localização dos salões. Entre os 100 entrevistados, 67% (67/100) mencionaram espontaneamente que conseguiram completar a tarefa de incluir a localização rapidamente, e elogiaram o fluxo escolhido. Apenas 1,4% (1) reconheceu que conseguiu completar a proposta, mas não fez nenhum elogio. No entanto, 2,98% (2) sentiram dificuldades na realização da tarefa, e seus comentários foram classificados como “negativos” (**Figura 8**).

As dificuldades foram mais proeminentes na tarefa de incluir o horário comercial. Entre os participantes, 84% (84/100) relataram uma experiência “muito positiva”, 1% (1/100) relatou uma experiência “positiva”, 4% (4/100) tiveram uma experiência “negativa” e 11% (11/100) tiveram uma experiência “muito negativa”. Entre os que tiveram dificuldades (12), 75% (9/12) eram brasileiros e 25% (3/12) eram portugueses. Alguns comentários dos participantes brasileiros incluíram: “Não ficou claro se podíamos editar a hora introduzindo números ou se havia opções de hora pré-estabelecidas” e “Acho que devia haver um botão ou uma seta para indicar que é possível personalizar a hora”. Por outro lado, os participantes portugueses referiram a falta de clareza sobre qual o botão que deve ser ativado para personalizar a hora.

Os aspetos de “acessibilidade” e “design gráfico” voltaram a surgir, com sugestões de melhoria relacionadas com o tamanho da letra e a atratividade, respetivamente.

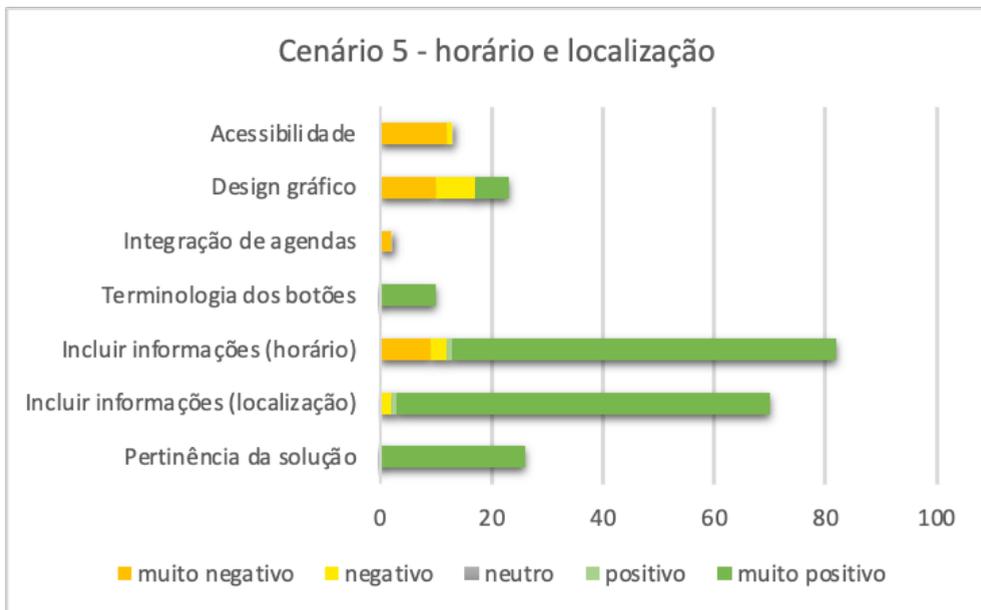


Figura 8 - Percepções sobre a inclusão do horário e da localização dos salões.

5.6. Informações sobre o horário e os dias de funcionamento (cenário 6)

Nesta tarefa, os participantes foram convidados a preencher informações sobre os seus clientes numa área especificamente destinada a armazenar estes dados. Do total de entrevistados, 63% (63/100) consideraram que se tratava de uma excelente opção para localizar mais facilmente os dados (**Figura 9**).

Foram feitos 80 comentários sobre a área de informação ao cliente, sendo 41% (33/80) classificados como “positivos”, 7,5% (6/80) como “negativos” e 51% (41) como “muito negativos”. Do total de opiniões negativas (47), 89% (42/47) foram expressas por

cidadãos brasileiros. A principal crítica foi a falta de pormenor na informação aos clientes. Alguns exemplos de comentários recolhidos sobre este tema são: “Acho que a secção de informações está incompleta. Deveria ser mais detalhada”, “ainda faltam informações importantes, como estilo de corte, cor e cabeleireiro preferido”, “falta uma quantidade significativa de informações relevantes, como data de aniversário e cortes anteriores” e “faltam tantas informações que acho que os dados disponíveis poderiam ser mais detalhados”. Conforme observado na **Figura 1**, essa área teve a menor perceção entre os entrevistados, principalmente entre os participantes brasileiros.

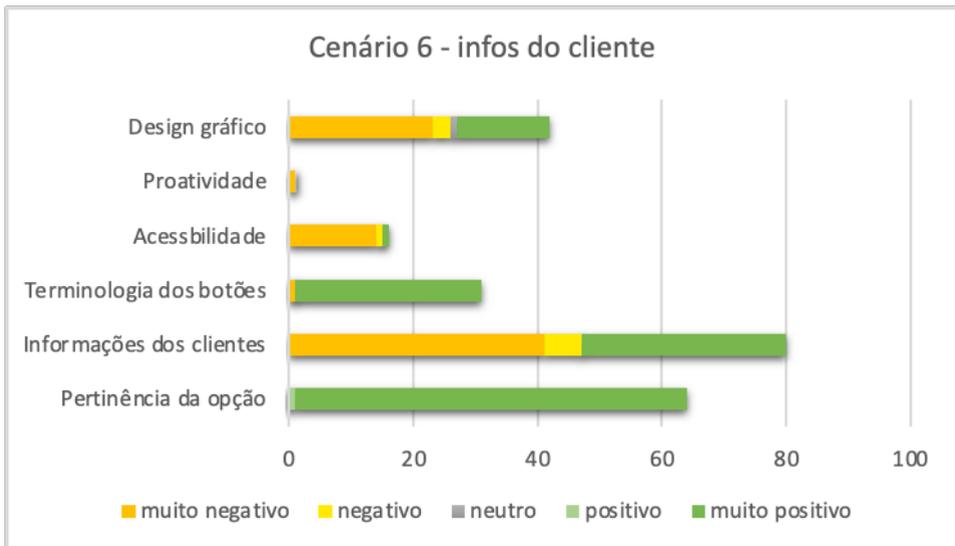


Figura 9 – Percepções sobre a área com os dados dos clientes.

Em relação aos aspetos “acessibilidade” e “design gráfico”, houve críticas quanto ao pouco espaço disponibilizado para os dados. Quarenta por cento dos participantes mencionaram que a área era muito pequena para a quantidade de informações que precisavam ser inseridas, resultando em consequências significativas para ambas as categorias.

6. Conclusões

A análise de UX descrita neste artigo empregou uma combinação de técnicas para avaliar tanto as qualidades hedônicas quanto pragmáticas do produto. Essas técnicas incluíram a avaliação de identificação, interesse, incentivo e atratividade. Todo o processo, incluindo seus elementos e etapas, foi detalhado no artigo. Um dos principais componentes da avaliação envolveu os participantes classificando cada cenário proposto em uma escala ordinal de 1 a 6. Além disso, uma análise abrangente de conteúdo dos comentários espontâneos dos participantes durante os testes de experiência de UX foi realizada e discutida.

Estatísticas descritivas foram utilizadas pelos pesquisadores para processar as avaliações de experiência dos participantes, que representaram dados qualitativos. A mediana, primeiro quartil e terceiro quartil, que são adequados para escalas ordinais, foram usados para fornecer uma opinião holística da amostra do estudo para cada cenário. Gráficos de radar foram utilizados para aumentar a perspetiva geral e resumir as opiniões dos entrevistados, destacando as divergências entre as populações portuguesa e brasileira.

A análise dos comentários dos participantes foi essencial para entender as avaliações atribuídas a cada cenário. Isso permitiu identificar áreas que não foram bem aceitas ou que poderiam ser melhoradas. A sistematização dos comentários levou à identificação de 28 sugestões de melhoria, incluindo renomear botões, abordar questões de acessibilidade (12 sugestões), integração com plataformas (6), confirmações via SMS (4) e informações detalhadas sobre os clientes (6).

Além disso, para medir a usabilidade geral do protótipo conceitual, todos os utilizadores foram convidados a completar um questionário SUS. Os resultados indicaram que a solução proposta foi útil e relativamente fácil de usar, embora algumas inconsistências tenham sido identificadas.

No geral, o método de P&D apresentado neste artigo mostrou-se eficiente na análise da experiência do utilizador, resultando em melhorias diretas na interface do sistema de conversação e na configuração de dados. Além disso, as diversas técnicas de coleta de dados empregadas ajudaram a identificar diferenças percebidas entre os dois grupos: cidadãos brasileiros e portugueses, em termos da configuração da interface do sistema de conversação.

Quanto à comparação entre os grupos, foi possível identificar um conjunto de diferenças na forma como avaliaram a plataforma. Isto é uma indicação importante de que diferenças culturais podem interferir na experiência do utilizador de um sistema. Por exemplo, uma parte significativa dos brasileiros não considerou a terminologia dos botões suficientemente clara. Em geral, eles acharam que os nomes escolhidos dificultam a localização de informações, o que não foi o caso dos cidadãos portugueses. Este comportamento se deve, em grande parte, às diferenças entre as palavras utilizadas no português do Brasil e no português de Portugal. No geral, os cidadãos brasileiros reagiram menos positivamente a todos os cenários, especialmente ao cenário 6. Uma possível razão para essa diferença, de acordo com 92% (46/100) dos entrevistados brasileiros, é que agendar compromissos por telefone deixou de ser o método mais comum usado em salões de beleza. A ferramenta mais frequentemente adotada é o WhatsApp. Portanto, este é outro exemplo importante de como diferenças culturais podem impactar a percepção dos utilizadores sobre sistemas e plataformas. Daí a necessidade de uma amostra diversificada para considerar a diversidade de opiniões em relação a novos produtos e soluções.

Agradecimentos. Este trabalho é apoiado pelo Fundo Europeu de Desenvolvimento Regional (FEDER), por meio do Programa Operacional de Competitividade e Internacionalização (COMPETE 2020) do quadro Portugal 2020 [Projeto CallBob com Número 038501 (POCI-01-0247-FEDER-038501)] e também é apoiado financeiramente por fundos nacionais por meio da FCT - Fundação para a Ciência e Tecnologia, I.P., no âmbito do projeto UIDB/05460/2020.

Referências

- Abreu, J., Camargo, J., Santos, R., Almeida, P., Beça, P., & Silva, T. (2021). UX Evaluation Methodology for iTV: Assessing a Natural Language Interaction System. In *Communications in Computer and Information Science* (Vol. 1433). https://doi.org/10.1007/978-3-030-81996-5_11
- Bernhaupt, R., & Pirker, M. M. (2014). User interface guidelines for the control of interactive television systems via smart phone applications. *Behaviour and Information Technology*. <https://doi.org/10.1080/0144929X.2013.810782>
- Brüggemeier, B., Breiter, M., Kurz, M., & Schiwy, J. (2020). User Experience of Alexa when controlling music: Comparison of face and construct validity of four questionnaires. *ACM International Conference Proceeding Series*, (pp. 1-9). <https://doi.org/10.1145/3405755.3406122>
- Guerino, G. C., & Valentim, N. M. C. (2020). Usability and User eXperience Evaluation of Conversational Systems: A Systematic Mapping Study. *ACM International Conference Proceeding Series*, (pp. 427-436). <https://doi.org/10.1145/3422392.3422421>
- Iniesto, F., Coughlan, T., & Lister, K. (2021). Implementing an accessible conversational user interface: Applying feedback from university students and disability support advisors. *Proceedings of the 18th International Web for All Conference, W4A 2021*, (pp. 1-5). <https://doi.org/10.1145/3430263.3452431>
- International Organization for Standardization. (2010). ISO 9241-210: Ergonomics of human-system interaction - Human-centred design for interactive systems. In *International Organization for Standardization*.
- Kiseleva, J., Williams, K., Jiang, J., Awadallah, A. H., Crook, A. C., Zitouni, I., & Anastasakos, T. (2016). Understanding user satisfaction with intelligent assistants. *CHIIR 2016 - Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval*, (pp.121-130). <https://doi.org/10.1145/2854946.2854961>
- Lallemant, C., & Koenig, V. (2020). Measuring the Contextual Dimension of User Experience: Development of the User Experience Context Scale (UXCS). *ACM International Conference Proceeding Series*, (pp. 1-13). <https://doi.org/10.1145/3419249.3420156>
- Le Pailleur, F., Huang, B., Léger, P. M., & Sénécal, S. (2020). A New Approach to Measure User Experience with Voice-Controlled Intelligent Assistants: A Pilot Study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-49062-1_13
- Ludwig, R., Bachmann, A., Buchholz, S., Ganser, K., Glänzer, D., & Matarage, A. (2021). How to measure ux and usability in today's connected vehicles. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-030-55307-4_3

- Pettersson, I., Lachner, F., Frison, A. K., Riener, A., & Butz, A. (2018). A bermuda triangle? - A review of method application and triangulation in user experience evaluation. *Conference on Human Factors in Computing Systems – Proceedings*, (pp. 1-16). <https://doi.org/10.1145/3173574.3174035>
- Trippas, J. R., Spina, D., Thomas, P., Sanderson, M., Joho, H., & Cavedon, L. (2020). Towards a model for spoken conversational search. *Information Processing and Management*, 57(2), 102162. <https://doi.org/10.1016/j.ipm.2019.102162>
- Wattarachchi, W. D., Hewagamage, K. P., & Hettiarachchi, E. (2022). Emotional Keyboard: To Provide Adaptive Functionalities Based on the Current User Emotion and the Context. *Interaction Design and Architecture(s) Journal*, (54), 147 – 174. <https://doi.org/10.55612/s-5002-054-007>

