

Curadoria de chatbots: conceptualização, estratégias e indicadores de desempenho

Dário Reyes Reina¹, Clarice Cruz¹

dario.reyes.re@gmail.com; clarice.cruz@hop.digital

¹ Hop Research Group, Rua Líbero Leone 259, CEP 30493-145, Belo Horizonte, Minas Gerais, Brasil

DOI: [10.17013/risti.40.1-14](https://doi.org/10.17013/risti.40.1-14)

Resumo: Nos últimos anos o surgimento e expansão de inovações tecnológicas, como os chatbots, tem ocasionado a formação de novas áreas de desempenho e/ou novos perfis profissionais. Nesse sentido a presente pesquisa teve o objetivo de analisar o fenômeno da curadoria de chatbots, para o qual foram realizadas 10 entrevistas semiestruturadas a profissionais experientes no desenvolvimento e na curadoria de chatbots. Os resultados foram organizados em três apartados: 1) a conceptualização da curadoria de chatbots, 2) estratégias utilizadas na curadoria e 3) indicadores de desempenho. Por último, as conclusões finais sintetiza os principais achados do estudo, discute-se sobre a relevância da curadoria de chatbots e compartilha possíveis linhas de pesquisa.

Palavras-chave: curadoria de chatbots; teste de chatbots; agentes conversacionais

Chatbots curation: conceptualization, strategies and performance indicators

Abstract: In recent years, the emergence and expansion of technological innovations, such as chatbots, has led to the formation of new work areas of jobs and / or new professional profiles. In this sense, the present research aimed to analyze the phenomenon of chatbots curation, for which 10 semi-structured interviews were conducted with professionals experienced in the development and chatbots curatorship. The results were organized in three sections: 1) the conceptualization of chatbots curation, 2) strategies used and 3) performance indicators. Finally, the main conclusions of the study are summarize, it is discuss the relevance of chatbots improvement and possible lines of research.

Keywords: chatbots curation; Chatbots testing; conversational agents

1. Introdução

A recente popularização dos chatbots, ao igual que outros tipos de agentes conversacionais, têm causado no mercado de trabalho o surgimento de novos cargos e/ou de novas tarefas relacionadas com o desenvolvimento dessas soluções tecnológicas. O que começou com tímidas aproximações de profissionais de diferentes disciplinas como o design,

o jornalismo, a linguística, a engenharia de sistemas, a ciência da computação, o marketing, entre outras, que aprenderam na prática o que funcionava e o que não, solidificou-se ao longo do tempo em um conjunto de conhecimentos e práticas muito especializadas.

No exercício profissional na área voltou-se normal falar de especialistas em desenho de conversas e fluxos de diálogos, analistas de interação, criação de “chatbot personas”, estruturação de bases de conhecimento ou mapeamento de intenções dos usuários, por mencionar alguns exemplos. Um lugar de destaque das discussões contemporâneas trata-se da conveniência da maturação gradativa dos chatbots num processo nomeado de “curadoria de chatbots”.

Contudo, chama atenção a pouca consolidação da evidência relacionada com essa nova atividade. Desse modo, a presente pesquisa teve como objetivo analisar em que consiste a curadoria de chatbots, as estratégias e indicadores de desempenho que usam os trabalhadores dedicados ao desenvolvimento e aprimoramento dessas tecnologias. Para isso, realizou-se uma pesquisa qualitativa através de entrevistas semi-estruturadas aplicadas a 10 profissionais dedicados ao desenvolvimento e a curadoria de chatbots. As entrevistas foram transcritas e analisadas por meio da análise de conteúdo temático segundo Bardin (Bardin, 2011)

Começa-se com uma revisão de literatura que brevemente apresenta as linhas de pesquisa recentes sobre os chatbots, seu desenho, indicadores de desempenho e as propostas para sua avaliação.

Na seguinte seção, apresenta-se mais detalhadamente a metodologia de pesquisa que orientou o estudo e os participantes do mesmo. Posteriormente, apresentam-se os resultados divididos em três partes: 1) conceptualização de curadoria de chatbots, 2) estratégias implementadas na curadoria e 3) os principais indicadores de desempenho utilizados. Por último, encontram-se as conclusões finais do estudo e compartilha-se possíveis linhas de pesquisa.

2. Revisão de literature

Os chatbots são softwares que interagem com as pessoas usando linguagem natural (Dale, 2016), simulando as conversas que temos entre humanos (Ciechanowski, Przegalinska, & Wegner, 2018; Morrissey & Kirakowski, 2013) e usando, principalmente, meios escritos.

Nos últimos anos devido à difusão dos chatbots em múltiplos e diversos domínios tem havido um crescente volume de estudos que focalizam sua atenção neles. No entanto, em particular a literatura sobre curadoria de chatbots é pouca, sendo que a maioria dos trabalhos abordam áreas correlatas.

A maior parte das pesquisas tem centrado seus esforços no entendimento dos fatores que influenciam a qualidade da interação humano-chatbot. Um caso sobressalente é a revisão de literatura de Chaves (2019), que sintetizou a evidência de 58 artigos em três grandes categorias de traços que deveriam ter os chatbots para satisfazer as expectativas dos usuários: 1) inteligência conversacional, que se refere às características dos chatbots que ajudam na gestão da conversa para que seja fluida e pertinente, 2) inteligência social, relacionada com as capacidades dos chatbots para entender e

adequar-se às convenções socialmente aceites, e 3) a personificação, que se trata da influência na interação humano-chatbot da atribuição de traços humanos aos chatbots.

Na área de indicadores de desempenho e modelos de avaliação de chatbots, pesquisadores como Radziwill (2017), Venkatesh (2018) e Kuligowska (2015), tem proposto um conjunto de atributos chaves que abrangem uma variedade de características que permitiriam uma avaliação integral do funcionamento desta tecnologia (ver Tabela 1). Estes atributos e indicadores são significativos na curadoria de chatbots, pois dirigem o olhar para os fatores mais relevantes, geram padrões de boas práticas e ajudam a estabelecer pontos de referência para mensurar se efetivamente as mudanças feitas no chatbot garantem uma boa experiência do usuário.

Outra abordagem seguida por pesquisadores como Vasconcelos (2017) e Bozic (2019), relaciona-se com a criação de ferramentas que automatizam algumas atividades no teste de chatbots. Vasconcelos (2017) descreve uma ferramenta, Bottester, que baseado em um corpus prévio de dados sobre perguntas e respostas frequentes simula a interação dos usuários com o chatbot. O objetivo desta solução é ajudar no aprimoramento do chatbot até que seja o suficientemente robusto, sem problemas críticos, para que possa ser liberado para testes com usuários reais.

Radziwill (2017)	Venkatesh (2018)	Kuligowska (2015) *
<p>Desempenho</p> <ul style="list-style-type: none"> Robustez para entrada inesperada (% de sucessos) Fornecer canais de escalação apropriados (% de sucessos) <p>Humanidade</p> <ul style="list-style-type: none"> Transparente (% de usuários que o classificaram corretamente como chatbot) Capaz de manter a discussão temática (pontuação de usuário ou de testador de 0 a 100) Capaz de responder a perguntas específicas (% de sucessos) <p>Cordialidade</p> <ul style="list-style-type: none"> Fornecer saudações e personalidade agradável (pontuação do usuário ou do testador de 0 a 100) Divertido, envolvente (pontuação do usuário ou do testador de 0 a 100) <p>Acessibilidade</p> <ul style="list-style-type: none"> Pode detectar significado e intenção (% de sucessos) Responde adequadamente às sugestões sociais (% de sucessos) 	<p>Experiência do usuário com a conversação</p> <ul style="list-style-type: none"> Avaliação dos usuários Avaliação dos usuários frequentes <p>Envolvimento</p> <ul style="list-style-type: none"> Número de turnos no diálogo Duração total da conversa <p>Coerência</p> <ul style="list-style-type: none"> Taxa de erro de resposta: Número de respostas incoerentes / Número total de enunciados <p>Domínio de temas</p> <ul style="list-style-type: none"> Medida de entropia (grau de aleatoriedade) na distribuição de conversas em diferentes temas. <p>Profundidade da Conversação</p> <ul style="list-style-type: none"> Média de turnos consecutivos no mesmo tópico <p>Diversidade tópica / amplitude de conversação</p> <ul style="list-style-type: none"> Tamanho do vocabulário Diversidade tópica 	<ul style="list-style-type: none"> Aparência Visual Forma de implementação no site Sistema de síntese da fala Conhecimento de temas básicos Conhecimento de temas especializados Apresentação de capacidades e funcionalidades adicionais Habilidades de conversação Sensibilidade ao contexto Traços de personalidade Opções de personalização Respostas de emergência em situações inesperadas Possibilidade de avaliar o chatbot e o site pelo usuário <p>*Os atributos são avaliados de 1 a 5 segundo uma série de características</p>

Tabela 1 – Atributos para avaliação de chatbots

Por sua parte Bozic (2019), propõe o que chama “Teste de chatbot baseado em planejamento” (*Planning-based chatbot testing*) para avaliar de forma automatizada a capacidade do chatbot para atingir um objetivo, suportado no mapeamento prévio de um conjunto de condições e de ações que devem ser cumpridas. Assim, no seu estudo de caso sobre um chatbot com o intuito de guiar a reserva em um hotel, os pesquisadores geraram múltiplas sequências de mensagens que simulavam as possíveis interações dos usuários e testaram automaticamente se as mesmas conseguiram atingir o objetivo de reservar.

Resumindo, tanto Vanconcelos (2017) como Bozic (2019) propõem contextos experimentais sob controle dos pesquisadores com intuito de fazer testes, e, em consequência, tomar as medidas necessárias para robustecer os chatbots *antes* de serem liberados aos usuários finais.

3. Metodologia

Realizou-se uma pesquisa qualitativa por meio de entrevistas semi-estruturadas aplicadas a profissionais brasileiros com experiência no desenvolvimento e na curadoria de chatbots. No total foram realizadas 10 entrevistas considerando o princípio de saturação teórica, segundo o qual, a coleta de dados é interrompida “quando se constata que elementos novos para subsidiar a teorização almejada (ou possível naquelas circunstâncias) não são mais apreendidos a partir do campo de observação” (Fontanella et al., 2011, p. 1).

Como critério único de inclusão os participantes deviam ter no mínimo um ano de bagagem trabalhando no desenvolvimento e na curadoria de chatbots, isto com o objetivo de garantir a pertinência da informação ao se tratar de pessoas com experiências e conhecimentos significativos (ver Tabela 2).

Participante	Anos de experiência com chatbots e outros agentes conversacionais	Gênero
1	1 ano	Masculino
2	2 anos	Masculino
3	4 anos	Masculino
4	4 anos	Masculino
5	4 anos	Feminino
6	4 anos	Feminino
7	3 anos	Feminino
8	8 anos	Feminino
9	5 anos	Feminino
10	1 ano	Feminino

Tabela 2 – Pessoas entrevistadas

As entrevistas, prévio consentimento dos participantes, foram realizadas virtualmente entre fevereiro e abril de 2020, para posteriormente serem transcritas e analisadas

por meio da análise de conteúdo temático segundo Bardin (Bardin, 2011). O roteiro da entrevista abordou três temas, em primeiro lugar, as características fundamentais da curadoria de chatbots, em segundo lugar as principais estratégias usadas na curadoria, e no terceiro os indicadores de desempenho usados nesta atividade. Nesta última seção, perguntou-se espontaneamente aos participantes quais eram os indicadores que usavam no seu cotidiano de trabalho, o motivo de uso, suas vantagens, desvantagens e caso fosse possível, que compartilhassem um caso ou exemplo real de uso. Foram coletados todos os indicadores mencionados pelos participantes e posteriormente foram agrupados em categorias maiores segundo suas similaridades e diferenças.

A análise foi feita em três fases: 1) pré-análise: consiste na familiarização com os dados e a geração das primeiras categorizações; 2) exploração do material: trata-se do aprofundamento nos dados e a criação de categorias maiores; e 3) tratamento dos resultados, inferência e interpretação: gerando conclusões e modelos explicativos, e analisando as relações entre categorias e sua concordância ou não com a literatura prévia. Em particular, os resultados foram sintetizados gerando uma definição emergente de curadoria de chatbots, uma figura para mapear as estratégias usadas na curadoria e uma tabela com a classificação dos indicadores de desempenho citado pelos participantes.

4. Resultados

4.1. Conceptualização da curadoria de chatbots

Para os participantes a curadoria de chatbots tem três características fundamentais:

I) Análise das interações reais dos usuários: a curadoria é uma atividade que começa uma vez que o chatbot já está sendo utilizado pelos usuários finais. Nesse sentido, os testes de qualidade e usabilidade embora sejam recomendados no processo de desenvolvimento, trata-se de atividades essencialmente diferentes da curadoria já que acontecem antes dos chatbots serem liberados totalmente aos usuários.

Rosenberg (2006) propõe o conceito de “aprendizado pelo uso” para se referir ao processo de aprendizado que só tem início depois que a nova tecnologia começa a ser utilizada pelos usuários finais. Segundo o pesquisador, há algumas características dos artefatos tecnológicos impossíveis de serem previstas e que só revelar-se-iam depois de sua utilização intensiva e prolongada. Este aprendizado seria acumulativo e paulatino, gerando pequenos ganhos de eficiência que somados constituem um impacto muito maior no desempenho da tecnologia.

A curadoria de chatbots é um exemplo claro de aprendizado pelo uso. Toda vez que a partir da análise das interações dos usuários identificam-se imprevistos ou novos requerimentos que deveriam ser integrados ao chatbot para garantir sua qualidade.

Essa análise da interação é feita de duas formas que se retroalimentam:

Análise Vertical: consiste no entendimento do fluxo de interação dos usuários com o chatbot, qual caminho foi percorrido ou sequência de interações que tiveram com a máquina. Por exemplo, caso o chatbot tivesse o objetivo de auxiliar na compra de uma passagem, a análise vertical ajuda a identificar em que ponto do processo de reserva os usuários estão tendo dificuldades.

Análise Horizontal: abordagem abrangente sobre a interação dos usuários com o chatbot, analisando as perguntas, assuntos ou requerimentos mais comuns e a capacidade da máquina para resolvê-las.

II) Processo paulatino e contínuo de aprimoramento: a curadoria segue como princípio norteador a metáfora de um ser vivo, nesse caso o chatbot, que gradualmente vai crescendo, maturando e adquirindo maiores capacidades para lidar com múltiplos cenários e requerimentos mais complexos.

Este crescimento paulatino e contínuo é devido a vários motivos: 1) a medida que os usuários interagem com os chatbots e conhecem suas capacidades começam a exigí-lhes mais; 2) surgimento de novas necessidades dentro das organizações, por exemplo, lançamento de novos produtos ou novas regulamentações, que requerem a atualização do chatbot; 3) requerimentos dos usuários que não tinham sido mapeados durante o processo inicial de desenvolvimento.

É importante ressaltar que esse aprimoramento do chatbot não acontece de maneira automática, pelo contrário requer a dedicação de uma ou várias pessoas, os encarregados da curadoria, que devem julgar a pertinência e custo benefício das mudanças no chatbot.

III) Arelado aos objetivos do chatbot: a curadoria está intimamente relacionada com os objetivos que esperam ser atingidos com o chatbot, ou seja, trata-se de uma atividade situada no contexto e na estratégia definida por cada organização. Alguns exemplos compartilhados pelos participantes podem deixá-lo mais claro: caso o chatbot tivesse o objetivo de diminuir o número de pessoas procurando atenção telefônica, a curadoria provavelmente vai estar focada em garantir a resolução das dúvidas dos usuários pelo chat e na análise da retenção desse canal. Ou se o chatbot tivesse o objetivo de qualificar leads e se enquadra dentro de uma estratégia comercial, a curadoria vai estar focada no engajamento e a análise da qualidade da informação coletada dos leads ao longo das interações com o chatbot.

Esta característica da curadoria de chatbots é singular porque vai além do estritamente relacionado com o funcionamento da tecnologia, o curador deve visar um alinhamento entre os objetivos, a estratégia da empresa, as capacidades do chatbot e as características dos usuários.

Em síntese, considerando as três características identificadas define-se curadoria de chatbots como o processo de aprimoramento paulatino e contínuo baseado na análise da interação dos usuários com eles, particularmente, examinando sua capacidade de responder aos objetivos previamente definidos.

4.2. Estratégias na curadoria de chatbots

Na realização da curadoria de chatbots há duas grandes estratégias, o foco na identificação e conserto de problemas e o foco na evolução dos chatbots (Figura 1). No primeiro caso, estamos falando da ênfase na identificação de falhas ou vazios que devem ser solucionados por se tratarem de requerimentos dos usuários que previamente não tinham sido mapeadas no processo de desenvolvimento.

No segundo caso, trata-se de aprimoramentos emergentes, ou seja, novas características que o chatbot irá ganhar. Neste caso não se trata estritamente de “erros do chatbot” e sim de oportunidades para melhorar suas capacidades além do que inicialmente tinha sido planejado e implementado.

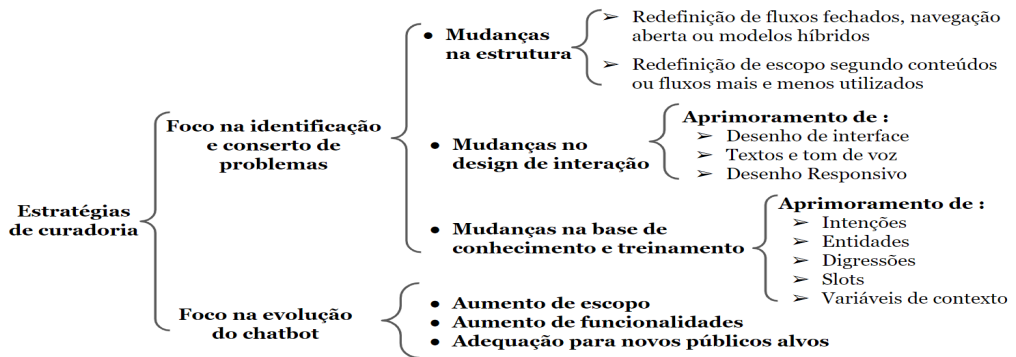


Figura 1 – Estratégias usadas na curadoria de chatbots

Foco na identificação e conserto de problemas:

- **Mudanças na estrutura:** geralmente os chatbots se enquadram em 3 tipos de estrutura. Em um extremo estão aqueles que oferecem exclusivamente fluxos de interação fechados, nos quais se espera que os usuários naveguem usando, principalmente, botões, menus, ou perguntas predefinidas que são sugeridas. No outro extremo, estão os chatbots que oferecem uma interação totalmente aberta, permitindo que em todo momento os usuários escrevam diretamente seus requerimentos no chat. No meio, encontram-se os chatbots com modelos híbridos, que misturam em determinados assuntos ou processos tanto a interação fechada como a possibilidade da interação livre pelo chat. Nesse contexto, parte da curadoria consiste em analisar as interações dos usuários e determinar as situações nas quais é melhor optar pelo fluxo fechado, a navegação aberta ou por modelos híbridos. Igualmente, uma mudança estrutural importante dos chatbots tem a ver com a análise dos fluxos, caso a estrutura for fechada, e os assuntos mais e menos consultados pelos usuários, caso for aberta, já que essa informação permite aos curadores decidir quais conteúdos e processos devem ser apagados ou ampliados.
- **Mudanças no design de interação:** um segundo ponto de atenção na curadoria está relacionado com a análise dos recursos visuais, textuais e técnicos usados na interface conversacional, bem como sua acessibilidade em diferentes plataformas. Pequenas mudanças na escrita de um texto, o tom de voz, ou nos recursos disponíveis no chatbot como botões, menus e carrosséis, terminam tendo um

impacto significativo na qualidade da interação. Em particular, os participantes mencionaram que testes A/B, análises heurísticas e o acompanhamento de alguns indicadores chaves, que aprofundaremos posteriormente, os ajudavam no julgamento desses elementos e na identificação das mudanças necessárias na interação.

- Mudanças na base de conhecimento e o treinamento do modelo inteligência artificial: parte rotineira da curadoria é o melhoramento da base de conhecimento e o treinamento do modelo de inteligência artificial por trás do funcionamento do chatbot. A maioria dos chatbots estão baseados no reconhecimento de intenções (intents) e entidades (entities), pelo que faz se necessário esclarecer esses conceitos. As intenções são propósitos ou objetivos dos usuários, por exemplo, em um chatbot de um banco uma intenção do usuário poderia ser “*saber o saldo da conta poupança*”. Enquanto as entidades, tratam-se de termos ou objetos, geralmente substantivos, que dão informação específica e relevante sobre uma intenção, no exemplo anterior “*conta poupança*” é uma entidade que permite saber a qual tipo de serviço financeiro o usuário está se referindo.

Deste modo, parte essencial da curadoria é analisar as escritas dos usuários e identificar nuances das intenções (utterances) e sinônimos ou novas entidades que deveriam nutrir a base de conhecimento do chatbot aumentando sua capacidade de entendimento.

Dependendo do chatbot, a curadoria também deverá prestar atenção ao que se conhece tecnicamente como Slots, digressões (digressions) e variáveis de contexto (context variables). Os slots são mecanismos que ajudam a coletar múltiplas informações compartilhados pelos usuários no percorrido do diálogo. No caso hipotético de um chatbot que ajuda a pedir pizza, um usuário poderia escrever no chat “*quero uma pizza de frango e de massa fina*”. Ao ter funcionando Slots, o chatbot poderia coletar a informação do requerimento de sabor do usuário (frango) e do tipo de massa (fina) e continuar perguntando por outros dados necessários para culminar o pedido da pizza.

As digressões referem-se às possibilidades do usuário voltar a um ponto anterior de seu diálogo com o chatbot. Retomando ao exemplo anterior, depois de ter pedido uma pizza de frango de massa fina o usuário poderia perguntar ao chatbot “*quanto demora em chegar o pedido?*”. Este responderia o tempo que demoraria em chegar, e depois por meio do funcionamento de digressões e slots poderia continuar o processo de pedido sem ter que coletar novamente as informações sobre o sabor da pizza e o tipo de massa.

Por último, estão as variáveis de contexto que ajudam a situar as circunstâncias e características particulares nas quais ocorre a conversação. Por exemplo, a informação sobre localização e hora são muito usadas como variáveis de contexto, já que situam a conversação permitindo ter um diálogo mais pertinente, obviamente, dependendo do foco particular do chatbot essas variáveis vão mudar.

Assim, a curadoria dedica esforços especiais na análise dos slots, nas digressões e nas variáveis de contexto já que são elementos que funcionando adequadamente permitem ter uma interação mais fluida e cumprem várias máximas do clássico princípio de cooperação das conversações (Price, 1975). Por um lado são recursos que ajudam a prevenir e corrigir erros, sendo coerente com a máxima da qualidade; situam a

conversação em um contexto com características particulares cumprindo a máxima da relevância; e permitem esclarecer possíveis ambiguidades e diminuir os desvios nas conversações satisfazendo a máxima do modo.

Foco na evolução dos chatbots:

O segundo grande foco na curadoria tem a ver com as estratégias de planejamento e desenvolvimento de novas características dos chatbots. Nesse caso, não se trata diretamente da resolução de problemas, e sim de oportunidades emergentes para aumentar sua abrangência, suas funcionalidades e suas capacidades para responder ao público alvo particular.

O aumento de escopo obedece, geralmente, a definição de novos objetivos que esperam ser alcançados pelo chatbot, fazendo necessário acrescentar assuntos sobre os quais o usuário pode perguntar ou novos processos que podem ser realizados no chatbot. Neste caso o desafio da curadoria além de avaliar a relevância desses acréscimos, está em criar e manter uma estrutura do chatbot o suficientemente flexível para aumentar seu escopo sem gerar conflitos com os desenvolvimentos anteriores e sem perder clareza e ordem.

O aumento de funcionalidades refere-se a utilização dos novos recursos que melhoram as capacidades do chatbot para entender e interagir com os usuários. Isto inclui integração com APIS para trazer e utilizar novas informações ao longo dos diálogos; inserção de novos modelos de processamento de linguagem natural, por exemplo, para facilitar a análise de sentimentos; disponibilização de novos recursos nas interfaces, por exemplo, menus, botões, checkbox, listas, carrosséis, sliders, imagens, vídeos, áudios, entre outros; além de qualquer outro desenvolvimento sob medida.

Para terminar, encontra-se a adequação do chatbot para novos públicos alvos, ou seja, a adaptação do chatbot para que a interação responda às particularidades de um grupo ou vários grupos de usuários através de mudanças no seu tom de voz, seu vocabulário e base de conhecimento. Por exemplo, caso um crescente uso de um chatbot por parte de pessoas de uma região específica seja identificado, o curador pode optar por passar de um “chatbot genérico” a um que ao perceber a localização do usuário utilize a mesma variação linguística das pessoas dessa região.

Sintetizando, a curadoria envolve um conjunto de estratégias complexas suportadas no entendimento do público alvo e sua interação com o chatbot, em certa medida, é como se os encarregados da curadoria cumprissem permanentemente o função fundamental de análise de requerimentos de usuários para garantir a qualidade do software (Barraza, 2017). O que demanda um olhar integral sobre múltiplos fatores: os objetivos do chatbot, o que motiva as pessoas a usar esse canal, quais assuntos procuram ou quais processos tentam realizar por ele, quais são as dificuldades que têm, e a análise do que funciona ou não no desenho conversacional (textos, fluxos, tom de voz, etc.) e nos recursos utilizados na interface (botões, carrosséis, etc.). Autores como Morales-Aguiar (2018) tem apontado o talento e os fatores humanos de ser elementos críticos no sucesso do desenvolvimento de software, neste caso, a curadoria se resulta em chatbots em processos paulatinos de aprimoramento, adquirindo maiores capacidades para entender ao usuário e se comunicar com fluidez.

4.3. Indicadores de desempenho usados na curadoria de chatbots

O monitoramento permanente de indicadores é uma tarefa fundamental da curadoria de chatbots, já que possibilita o julgamento do desempenho da solução e a avaliação do impacto das mudanças implementadas ao longo do processo.

Considerando que cada chatbot possui objetivos, características e públicos alvo particulares, os indicadores escolhidos na curadoria e sua interpretação vão depender de cada caso. Pelo mesmo motivo, os participantes manifestam a dificuldade da comparação dos indicadores entre diferentes chatbots, inclusive naqueles de um mesmo setor. Nesse sentido, a melhor alternativa para avaliar o desempenho é fazer um contraste permanente dos indicadores de cada chatbot versus seu próprio histórico.

Especificamente, foram identificados 4 tipos de indicadores usados rotineiramente na curadoria de chatbots (Tabela 3):

- *Indicadores de Cumprimento de Objetivos*: constitui o principal foco de atenção da curadoria, assim que na medida em que houver um melhoramento nesses indicadores o chatbot estaria apresentando um desempenho positivo. Estes indicadores variam em cada caso devido a que estão diretamente atrelados aos objetivos particulares definidos no momento do desenvolvimento do chatbot.

No entanto, geralmente são usados dois indicadores: I) Taxa de retenção, que se refere a % de usuários que resolve seu requerimento pelo chatbot, não precisando da utilização de outros canais de atenção.

II) Taxa de Conclusão, que se trata da % de usuários que consegue culminar com sucesso determinado processo ou objetivo dentro do chatbot. Por exemplo, se o objetivo do chatbot for qualificar leads por meio de 5 perguntas, a taxa de conclusão se refere a quantas pessoas chegaram até o fim e responderam às 5 perguntas. Além disso, é usual medir a “taxa de evasão ou análise de funil” em cada fase do processo. Voltando ao exemplo, seria calcular a % de leads que continuaram da pergunta 1 à pergunta 2, e assim sucessivamente.

- *Indicadores de Satisfação*: são indicadores que surgem ao pedir explicitamente ao usuário uma avaliação quantitativa de seu nível de satisfação da interação com o chatbot. Vale a pena salientar uma importante limitação desses indicadores que é a dificuldade das pessoas para julgar separadamente sua experiência com chatbot e da satisfação com a marca como um todo.

Embora exista uma variedade de formas para avaliar a satisfação, o Net Promoter Score (NPS), que usa uma nota de 1 a 10, e a avaliação por meio de 5 estrelas são as alternativas mais usadas. Em ambos casos, pela familiaridade dos usuários com as escalas e a possibilidade de fazer comparações com os níveis de satisfação de outros canais de atenção.

- *Indicadores de Assertividade*: trata-se de métricas que servem para avaliar a capacidade do chatbot para responder efetivamente os requerimentos dos usuários. Além disso, cumprem um papel fundamental na identificação de erros ou vazios que possui o chatbot. Nessa classificação encontram-se:

I) Gatilhos de Confusão: refere-se a % de interações que o chatbot não sabe como responder, pelo que deveria informar ao usuário: “isso não sei como responder”. Usualmente, a análise dos diálogos nos quais um gatilho de confusão é disparado termina sendo uma fonte direta para identificar limitações dos chatbots e planejar futuras soluções.

II) One Answer Success: trata-se da % de diálogos nos quais o chatbot teve a capacidade de resolver o requerimento do usuário precisando somente de uma interação. O indicador privilegia o princípio de economia na comunicação, no entanto, seu uso depende estritamente do desenho do chatbot, pois nem todos possuem essa lógica e propósito resolutivo.

III) Chatbot Rates: alude a % de respostas positivas que obteve o chatbot ao pedir ao usuário uma retroalimentação sobre sua resposta. Geralmente, o indicador é coletado perguntando textualmente ao usuário “Te ajudei? Sim ou Não” ou por meio de ícones de polegar para cima e para baixo.

Similar aos gatilhos de confusão, a análise dos casos que o usuário responde negativamente são uma fonte explícita para futuros aprimoramentos do chatbot. Contudo, o entendimento da situação problema deve ser realizado com muito cuidado já que podem acontecer falsos negativos, isto é, ocasiões nas quais o usuário disse que a resposta não o ajudou, mas o chatbot respondeu “corretamente” segundo seu prévio treinamento. Nessas circunstâncias, deve ser analisado o diálogo como um todo para ter uma melhor compreensão, e se for o caso, rever a base de conhecimento e as respostas estipuladas para cada intenção.

IV) Índice de Confiança das Intenções: usualmente os chatbots baseados em intenções permitem ter um índice de confiança do reconhecimento das mesmas em uma escala de 0 a 1, sendo 1 o máximo valor possível. A análise desse indicador permite a identificação de possíveis problemas na base de conhecimento e o treinamento do modelo de inteligência artificial, bem como ambiguidades entre diversas nuances (utterances) das intenções.

Tipo de Indicador	Indicador
Cumprimento de objetivos	<ul style="list-style-type: none"> - Taxa de retenção (% de usuários que resolve seu requerimento pelo chatbot, não precisando da utilização de outros canais de atenção) - Taxa de conclusão (% de usuários que conseguem culminar com sucesso determinado processo)
Satisfação	<ul style="list-style-type: none"> - Net Promoter Score (Avaliação de 1 a 10, sendo 10 a máxima nota) - 5 Estrelas (Avaliação de 1 a 5, sendo 5 a máxima nota)
Assertividade	<ul style="list-style-type: none"> - Gatilhos de confusão (% de interações que o chatbot não sabe como responder, pelo que deveria informar ao usuário: “isso não sei como responder”) - One Answer Success (% de diálogos nos quais o chatbot teve a capacidade de resolver o requerimento do usuário precisando só uma interação) - Chatbot Rates (% de respostas positivas quando o chatbot pede ao usuário por retroalimentação: “Te ajudei? Sim _ Não _”) - Índice de confiança das intenções

Tipo de Indicador	Indicador
Uso	<ul style="list-style-type: none"> - Usuários x período de tempo - Interações x período de tempo - Média de Interações por sessão - Tempo Médio por sessão - Intenções mais e menos usadas - Entidades mais e menos usadas - Assuntos mais o menos usados - Fluxos mais e menos usados

Tabela 3 – Indicadores usados na curadoria de chatbots

- *Indicadores de Uso*: são indicadores que ajudam a entender a utilização dos chatbots. Nessa classificação encontram-se algumas métricas genéricas como número de usuários e de interações por período de tempo, o tempo médio da interação do usuário com o chatbot, e a média de interações dos usuários, ou seja, quantas mensagens são intercambiadas em cada sessão.

Esses indicadores sempre devem ser analisados em contexto, levando em conta os objetivos do chatbot e a plataforma na qual está alojado. Por exemplo, um maior tempo médio da sessão de um chatbot em WhatsApp versus a mesmo chatbot alojado em um aplicativo poderia se explicar, simplesmente, porque os usuários de WhatsApp estão fazendo múltiplas atividades ao mesmo tempo, esquecendo de escrever no chatbot.

Por último, estão os indicadores sobre intenções, entidades, assuntos e fluxos mais e menos utilizados nas interações. Neste caso, trata-se de métricas que ajudam a determinar os principais interesses dos usuários e decidir quais conteúdos ou processos que podem ser realizados no chatbot deveriam ser ampliados, ou apagados.

Um achado relevante é a diferença entre os indicadores de desempenho utilizados no exercício profissional pelos encarregados da curadoria com aqueles mencionados pela literatura. Por um lado, Kuligowska (2015), Radziwill, (2017) e Venkatesh (2018) sugerem um amplo conjunto de atributos a serem considerados, que em grande medida, abrangem as características de inteligência conversacional, inteligência social e personificação mencionados por Chaves (2019) para oferecer uma boa interação humano-chatbot.

Por outro lado, no exercício profissional da curadoria ganha uma relevância muito maior os indicadores relacionados com o cumprimento de objetivos, que não foram identificados na bibliografia. Só implicitamente há alguns indicadores usados no cotidiano da curadoria com alguma relação com a literatura: os indicadores de satisfação são similares aos atributos experiência de usuário de Venkatesh (2018) e avaliação de usuário de Kuligowska (2015); os de assertividade com os atributos de desempenho e acessibilidade de Radziwill (2017) e os de habilidades de conversação e sensibilidade ao contexto de Kuligowska (2015); enquanto os indicadores de uso, tem relação com os atributos de envolvimento e profundidade da conversação de Venkatesh (2018).

5. Conclusões

A curadoria de chatbots é uma tarefa complexa que envolve um conjunto de habilidades, conhecimentos e estratégias altamente especializadas. Como foi apontado, é impossível prever todos os requerimentos dos usuários ou os cenários possíveis nos quais os chatbots vão ser usados. Nesse sentido, o aprimoramento contínuo e permanente por meio da curadoria é uma atividade fundamental para garantir a qualidade das interfaces conversacionais.

Embora existam avanços relevantes na área de aprendizado automático e no mercado se fala coloquialmente da capacidade das máquinas e dos algoritmos para aprender por si só, a presente pesquisa evidenciou uma realidade totalmente distinta. A curadoria está muito longe de ser automatizada, o julgamento, sensibilidade, e a capacidade de avaliar integralmente múltiplos fatores e indicadores para tomar decisões estratégicas continuam sendo atividades essencialmente humanas.

Uma das características fundamentais identificadas sobre a curadoria foi seu embasamento no aprendizado pelo uso. Este traço, é extensivo às pessoas dedicadas ao desenvolvimento e a curadoria de chatbots, já que partindo de sua própria experiência, seu conhecimento tácito, e a interação com outros colegas, têm conseguido consolidar estratégias para nortear seu trabalho profissional.

Uma importante limitação dos resultados da pesquisa foi a dificuldade para encontrar outros estudos para discutir. De fato, entre a pouca literatura correlata encontrada se fez evidente a grande diferença entre a bibliografia acadêmica e o exercício profissional, o caso mais palpável foram os indicadores de desempenho e os critérios de avaliação da qualidade dos chatbots, já que as propostas de autores como Radziwill (2017), Venkatesh (2018) e Kuligowska (2015) tinham pouca relação com as métricas efetivamente usadas na prática da curadoria.

Sugere-se a realização de novas pesquisas que aprofundem sobre o exercício da curadoria em outros contextos de atuação, provavelmente, setores intensivos no uso de novos desenvolvimentos de inteligência artificial empreguem pessoas para realizarem atividades análogas aqui analisadas. Igualmente, acredita-se que seja necessário aprofundar nas características que adquire a curadoria segundo as singulares de cada chatbot, particularmente, seria interessante entender as mudanças quando o chatbot atende um grande volume de interações sendo necessário contar com times dedicados e ferramentas que auxiliem o trabalho.

Referências

- Bardin, L. (2011). *Análise de Conteúdo*. São Paulo: Edições 70.
- Barraza, I.D., & Zepeda, V.V. (2017). Factores sociales y humanos que afectan el proceso de educación de requerimientos: una revisión sistemática. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (24),69-83. <https://doi.org/10.17013/risti.n.69-83>.
- Bozic, J., Tazl, O.A., & Wotawa, F. (2019). Chatbot Testing Using AI Planning. 2019 IEEE International Conference on Artificial Intelligence Testing (AITest). <https://doi.org/37-44.10.1109/AITest.2019.00-10>

- Ciechanowski, L., Przegalinska, A., & Wegner, K. (2017). The Necessity of New Paradigms in Measuring Human-Chatbot Interaction. https://doi.org/10.1007/978-3-319-60747-4_19
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811-817. <https://doi.org/10.1017/S1351324916000243>
- Fontanella, B.J, Luchesi, B.M., Saidel, M.G., Ricas, J.B., Turato, E.R., & Melo, D. G. (2011). Amostragem em pesquisas qualitativas: proposta de procedimentos para constatar saturação teórica. *Cadernos de Saúde Pública*, 27(2), 388-394. <https://dx.doi.org/10.1590/S0102-311X2011000200020>
- Morales-Aguiar, N., & Vega-Zepeda, V. (2018). Factores Humanos y la Mejora de Procesos de Software: Propuesta inicial de un catálogo que guíe su gestión. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (29), 30-42. <https://doi.org/10.17013/risti.29.30-42>.
- Morrissey, K., & Kirakowski, J. (2013). “Realness” in chatbots: establishing quantifiable criteria. *HCI*, 87-96. https://doi.org/10.1007/978-3-642-39330-3_10
- Kuligowska, K. (2015). Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents.1-16. <https://doi.org/10.18483/PCBR.22>
- Grice, H. P. (1975). *Logic and Conversation*. In *Logic and Conversation*. Leiden: Brill. doi: https://doi.org/10.1163/9789004368811_003
- Radziwill, N.M., & Benton, M.C. (2017). Evaluating Quality of Chatbots and Intelligent Conversational Agents. *ArXiv*, <https://arxiv.org/abs/1704.04579>.
- Rosenberg, N. (2006). Capítulo 6: O aprendizado pelo uso. Em N. Rosenberg, *Por Dentro da Caixa-preta: Tecnologia e Economia*. Campinas: Unicamp.
- Vasconcelos, M., Candello, H., Pinhanez, C.S., & Santos, T.D. (2017). Bottester: Testing Conversational Systems with Simulated Users. *IHC*, 2017, 1-4. <https://doi.org/10.1145/3160504.3160584>
- Venkatesh, A., Khatri, C., Ram, A., Guo, F., Gabriel, R., Nagar, A., Prasad, R., Cheng, M., Hedayatnia, B., Metallinou, A., Goel, R., Yang, S., & Raju, A. (2018). On Evaluating and Comparing Conversational Agents. *ArXiv*, <https://arxiv.org/abs/1801.03625>.