

# O que revelam os estudos secundários sobre chatbots na educação?

Leo Natan Paschoal<sup>1</sup>, Tayana Uchôa Conte<sup>2</sup>, Simone R. S. de Souza<sup>1</sup>

<sup>1</sup>Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo  
Avenida Trabalhador São-carlense, 400 - Centro – São Carlos – SP – Brasil

<sup>2</sup>Instituto de Computação – Universidade Federal do Amazonas  
Manaus, Brasil

paschoalln@usp.br, tayana@icomp.ufam.edu.br, srocio@icmc.usp.br

**Abstract.** *Chatbots are established with multiple intentions to support educational activities. The number of primary studies on the educational use of chatbots has been increasing. Consequently, secondary studies have emerged in state-of-the-art. However, there are no tertiary studies to provide a critical overview of the available literature. In particular, no understanding of what evidence has already been obtained by secondary studies and what still needs to be developed by the community. This paper seeks to alleviate this gap by categorising the secondary studies on educational chatbots. This paper offers an inventory of systematic mappings and systematic reviews. Hoping it will be helpful to researchers and practitioners interested in the field.*

**Resumo.** *Chatbots têm sido estabelecidos com múltiplos propósitos para apoiar atividades educacionais. Em razão disso, o número de relatos sobre o impacto decorrente do uso desse tipo de software tem aumentado e estudos secundários têm surgido na literatura. Apesar de existirem estudos secundários, ainda não há estudos terciários que ofereçam uma visão crítica sobre a literatura disponível, especialmente não há um entendimento concreto sobre as evidências já obtidas e o que ainda é preciso ser desenvolvido pela comunidade. Este artigo busca amenizar essa lacuna, categorizando os estudos secundários relacionados a chatbots educacionais e oferecendo uma análise crítica sobre os estudos realizados. Um inventário de mapeamentos e revisões sistemáticas é fornecido para pesquisadores e profissionais que possuem interesse na temática.*

## 1. Introdução

Chatbots são sistemas interativos que se comunicam com os usuários por meio de linguagem natural [AbuShawar and Atwell 2015, Allouch et al. 2021]. Esses sistemas têm sido incluídos em ambientes que carecem de interação e suporte, como os sistemas de comércio eletrônico e os sistemas de gestão da aprendizagem. Em particular, no contexto de sistemas de gestão da aprendizagem, os chatbots são estabelecidos para desempenhar diferentes tarefas, como resolver dúvidas dos estudantes sobre o curso [Ranoliya et al. 2017, Paschoal et al. 2022], estimular a participação do aluno em aula [Benotti et al. 2014], encorajar o estudante por meio de discurso motivacional [Ong and Soriano 2016], atuar como companheiro de conversação em um idioma do qual o usuário não é nativo [Ayedoun et al. 2015], dentre outros.

As tarefas realizadas pelo chatbot dependem do papel que ele desempenha no sistema educacional. Tais tarefas podem ser realizadas por iniciativa do próprio chatbot ou por solicitação do aluno. O chatbot irá conduzi-las a partir de uma interação, feita por meio de voz ou texto [Maicher et al. 2017]. Essa interação tende a ser realizada sem restrições linguísticas [AbuShawar and Atwell 2015]. Nesse sentido, espera-se que

durante a interação, o aluno possa escrever ou falar com chatbot livremente, sem precisar selecionar um conjunto de opções previamente definidas. Ainda, de modo complementar, a interação pode envolver modalidades de comunicação não verbal (*e.g.*, gestos, expressões faciais ou corporais), que exigem do chatbot uma representação facial ou corporal [Santos-Perez et al. 2013, Wik and Hjalmarsson 2009].

Essas diferentes modalidades de interação possibilitam um amplo conjunto de perspectivas de uso. Diante disso, pesquisadores têm manifestado interesse em explorar os diferentes usos dos chatbots em práticas e ambientes educacionais [Bezverhny et al. 2020]. Isso é possível perceber dada a quantidade de estudos que reportam a concepção desses sistemas, em fóruns de divulgação científica na área. Há chatbots desenvolvidos para educação em saúde [Frangoudes et al. 2021], ensino de computação [Benotti et al. 2018, Paschoal et al. 2019], aprendizado de línguas [Jeon 2022], dentre outras. Existem estudos sobre chatbots para apoiar cursos massivos online [Pereira et al. 2019], sistemas tutores inteligentes [Nguyen et al. 2019] e mundos virtuais para educação [Herpich et al. 2016]. Além disso, nota-se que diferentes métodos de pesquisa têm sido adotados, como os estudos experimentais que buscam compreender os efeitos produzidos pelos chatbots em práticas educativas [Fryer et al. 2017].

Ao passo que estudos primários têm abordado os efeitos dos chatbots na educação, estudos secundários afloraram com a intenção de agregar as evidências existentes sobre as contribuições feitas no âmbito de chatbots educacionais. Nesse sentido, a comunidade científica tem demonstrado interesse em acumular as evidências de forma ordenada e precisa. Essas evidências têm sido reunidas por meio de métodos científicos consolidados, como os mapeamentos e as revisões sistemáticas da literatura [Kitchenham and Charters 2007, Petersen et al. 2015]. Tais métodos permitem fornecer novas interpretações sobre resultados alcançados pelos estudos primários, assim como aprimorar argumentos, conceitos e teorias [Brereton et al. 2007, Kitchenham et al. 2009]. Ademais, esses métodos de condução de estudos secundários oportunizam reconhecer questões que precisam de um maior aprofundamento e cuidado na área de pesquisa [Kitchenham et al. 2009, Kitchenham et al. 2010].

Há um entendimento que esses aglomerados de evidências são capazes de contribuir de diferentes formas com a pesquisa que está sendo desenvolvida na área. Apesar dos estudos secundários cobrirem diversos aspectos sobre os efeitos do uso dos chatbots na educação, até o momento não há um estudo terciário que analisa e agrupa os estudos secundários realizados. Por conta disso, o foco dos estudos secundários, suas questões norteadoras, lacunas e fragilidades ainda não foram discutidas. Desse modo, é importante identificar os estudos secundários sobre chatbots educacionais, indicando para a comunidade onde os estudos podem ser localizados, além de oferecer uma visão crítica e consolidada sobre quais evidências já foram reunidas e interpretadas, como também manifestar as evidências que ainda precisam ser reconhecidas.

Este artigo reporta uma pesquisa que busca amenizar a ausência de um estudo terciário. Em particular, o artigo parte de uma inquietação que almeja entender o que revelam os estudos secundários sobre chatbots educacionais. Este estudo reúne e analisa os estudos secundários sobre chatbots educacionais por meio da aplicação de diretrizes de [Kitchenham and Charters 2007]. Mapeamentos e revisões sistemáticas que adotaram processos consolidados para planejamento e condução de estudos secundários foram agrupados. Um conjunto de estudos secundários foi localizado e as descobertas desses estudos foram relacionadas buscando oferecer evidências para esclarecer a inquietação.

## 2. Materiais e métodos

Para realização deste estudo, o método de condução de revisões sistemática da literatura reportado por [Kitchenham and Charters 2007] foi adotado. Isso foi feito porque um estudo terciário pode ser compreendido como uma revisão sistemática de estudos secundários. Nesse sentido, atividades foram idealizadas e realizadas. Nesta seção, as etapas de planejamento e condução do estudo terciário serão descritas.

## 2.1. Planejamento

O estudo terciário tem início a partir da inquietação “*o que revelam os estudos secundários sobre chatbots na educação?*”. Para respondê-la e desenvolver um debate que agregue valor para a área, quatro questões de pesquisa foram estabelecidas. Cada uma das questões de pesquisa é indicada na Tabela 1, juntamente com os motivos que estimularam o surgimento de cada uma delas. Espera-se que ao responder essas questões, seja possível categorizar e oferecer uma análise crítica sobre as informações contidas nos estudos secundários.

**Tabela 1. Questões de pesquisa e motivações do estudo terciário**

| ID  | Questão de pesquisa (QP)   | Motivação   |
|-----|--|---|
| QP1 | Quais são os estudos secundários sobre chatbots educacionais?                                      | Essa questão busca caracterizar e apresentar os estudos secundários, indicando o local em que eles foram publicados, o ano da publicação e o período contemplado pelas buscas realizadas em cada estudo secundário. |
| QP2 | Quais são os escopos dos estudos secundários sobre chatbots educacionais?                          | Essa questão almeja identificar os assuntos que estão direcionando e sendo contemplados pelos estudos secundários.  |
| QP3 | Quais questões de pesquisa foram investigadas nos estudos secundários sobre chatbots educacionais? | Essa questão tem a finalidade de reunir as questões de pesquisa que já foram respondidas em estudos secundários.  |
| QP4 | Quais os desafios e as oportunidades de pesquisa para a temática de chatbots educacionais?         | Essa questão procura exteriorizar as oportunidades de pesquisa na área de chatbots educacionais e os desafios inerentes às pesquisas que estão sendo desenvolvidas.   |

A partir da especificação das questões de pesquisa, a estratégia para localizar os estudos secundários foi determinada. Inicialmente, adotou-se a recomendação de [Brereton et al. 2007] para fazer uma busca automatizada em base de dados que indexam artigos científicos publicados em diferentes veículos de divulgação científica. A partir disso, as fontes de busca foram escolhidas, a saber: ACM Digital Library<sup>1</sup>, IEEE Xplore<sup>2</sup>, Science Direct<sup>3</sup>, Scopus<sup>4</sup> e Web of Science<sup>5</sup>. Há três motivos que justificam a escolha dessas bases de dados: (i) elas catalogam estudos científicos de conferências e periódicos que possuem políticas editoriais rigorosas em termos de avaliação por pares; (ii) elas são recomendadas por pesquisadores que conduzem mapeamentos e revisões sistemáticas; (iii) os estudos sobre chatbots educacionais costumam ser publicados em eventos e periódicos que são indexados por essas bases de dados.

Mediante a definição das bases de dados, foi necessário organizar uma string de busca. Dois critérios foram adotados para apoiar a definição da string, são eles: população e intervenção [Kitchenham and Charters 2007]. A população remete a área de

<sup>1</sup>Mais informações disponíveis em: <https://dl.acm.org/>

<sup>2</sup>Mais informações disponíveis em: <https://ieeexplore.ieee.org/>

<sup>3</sup>Mais informações disponíveis em: <https://www.sciencedirect.com/>

<sup>4</sup>Mais informações disponíveis em: <https://www.scopus.com/>

<sup>5</sup>Mais informações disponíveis em: <https://www.webofscience.com/>

estudos da investigação, já a intervenção é a metodologia abordada por uma determinada questão. Nesse sentido, estabeleceu-se que a população é representada por “chatbot” e a intervenção por “revisão sistemática”. Variações dos termos que representam a população e a intervenção foram estipulados, conforme indicado na Figura 1. Elas delimitam as palavras-chave necessárias para compor a string de busca. Para essas palavras-chave, a string de pesquisa foi definida como:

```
((`CHATBOT` OR `CONVERSATIONAL AGENT` OR `CHATTERBOT` OR `CHAT BOT` OR `DIALOGUE SYSTEM` OR `CONVERSATIONAL SYSTEM` OR `VIRTUAL AGENT` OR `VIRTUAL ASSISTANT`)) AND (`SYSTEMATIC LITERATURE REVIEW` OR `SLR` OR `SYSTEMATIC REVIEW` OR `SYSTEMATIC MAPPING` OR `MAPPING STUDY` OR `LITERATURE REVIEW`))
```

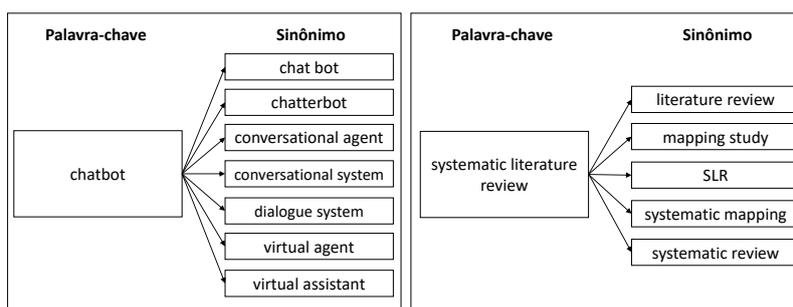


Figura 1. Termos que constituem a string de busca

Vale frisar que o termo “education” não foi incluído na string de busca e há uma justificativa para essa decisão. O termo e seus respectivos sinônimos poderiam restringir a busca pelos estudos. Então, a busca foi expandida para explorar os resultados de uma pesquisa em bases de dados que retornassem todos os possíveis estudos secundários sobre chatbots. Desse modo, foi necessário criar critérios que permitissem a inclusão somente de estudos secundários sobre chatbots educacionais. Critérios de seleção especificam condições para os estudos secundários serem incluídos ou excluídos no estudo terciário. Os critérios definidos para este estudo são apresentados na Tabela 2.

Tabela 2. Critérios utilizados na seleção dos estudos secundários

| Critérios de inclusão (CI)  | Critérios de exclusão (CE)  |
|---|---|
| CI-1: É um estudo secundário sobre chatbots na educação   | CE-1: O estudo não atende aos critérios de inclusão                     |
| CI-2: É um estudo revisado por pares  | CE-2: O texto do estudo secundário não está disponível                  |
| CI-3: É um estudo conduzido de modo sistemático, envolvendo definição de fontes de busca, questões de pesquisa e critérios de seleção dos estudos primários | CE-3: O estudo secundário está escrito em um idioma diferente do inglês |

Com o propósito de responder às questões de pesquisa abordadas neste estudo terciário, antes da etapa de condução do estudo terciário, uma tabela de apoio à extração de dados foi construída. Essa tabela é constituída por campos como título do estudo secundário, ano de publicação, período de publicação dos estudos primários, tipo de estudo (*i.e.*, revisão sistemática, mapeamento sistemático), objetivo de pesquisa, questões de pesquisa e oportunidades de pesquisa indicadas pelos estudos. Vale ressaltar que para classificar o tipo de estudo, optou-se por utilizar a terminologia apresentada pelos autores dos estudos secundários.

## 2.2. Condução

A primeira etapa da condução do estudo terciário consistiu na busca automática dos estudos secundários, utilizando as bases de dados supracitadas na seção anterior. Para tanto,

foi necessário adequar a string de busca para ser aceita pelos mecanismos de busca. Assim, cinco variações da string foram desenvolvidas - uma para cada fonte de busca, com cuidado para garantir que nenhum das palavras-chave fossem deixadas de lado ou recebessem uma importância maior que as demais. Os resultados das buscas foram exportados em formato “.bib” e organizados na plataforma Parsifal<sup>6</sup>. Essa plataforma de apoio à condução de revisões sistemáticas é capaz de identificar automaticamente estudos duplicados, ou seja, artigos disponíveis e indexados em mais de uma das bases de dados.

A segunda etapa da condução tem início após a eliminação de estudos duplicados. Ela focou na leitura dos títulos e resumos dos estudos não duplicados e na aplicação dos critérios de seleção retratados anteriormente. Estudos que atenderam a todos os critérios de inclusão e a nenhum dos critérios de exclusão foram encaminhados para a próxima etapa. Os estudos foram descartados quando satisfaziam um dos critérios de exclusão. Vale ressaltar que os estudos que não possuíam informações suficientes em seus títulos e resumos, foram encaminhados para a próxima etapa.

Durante a terceira etapa da condução, os estudos selecionados na etapa anterior foram lidos na íntegra, considerando aos critérios de seleção. Com base nessa leitura, o conjunto final de estudos foi definido. Posteriormente, os dados necessários para responder às questões de pesquisa foram extraídos.

Todas as etapas da condução foram realizadas no último trimestre de 2021. Durante este período, as buscas por estudos secundários foram feitas em cada uma das bases de dados e 234 artigos foram localizados. Vale salientar que nenhuma restrição foi adotada sobre o período em que os estudos secundários foram publicados. Com base nessas buscas, os metadados dos artigos (*i.e.*, título, resumo, URL do artigo, palavras-chave e número DOI - Digital Object Identifier) foram exportados das bases de dados e importados na ferramenta Parsifal. Na sequência, com apoio desta ferramenta, 82 estudos duplicados foram removidos, restando 152 artigos. Os títulos e resumos desses artigos foram lidos por um pesquisador e ele atribuiu um critério de seleção para cada artigo. Ao final dessa análise, nove artigos passaram para a etapa de leitura completa. O pesquisador fez a leitura completa dos nove artigos e ao aplicar os critérios de seleção, cinco estudos secundários foram selecionados. Posteriormente, os dados dos artigos foram extraídos. Visando responder às questões de pesquisa, uma análise foi realizada e apresentada na próxima seção.

### 3. Resultados e discussão

Esta seção descreve os resultados obtidos com o estudo terciário. Os resultados são mostrados com base nas questões de pesquisa definidas.

#### QP1 – Quais são os estudos secundários sobre chatbots educacionais?

A primeira questão do estudo terciário busca retratar os estudos secundários que empregaram procedimentos sistemáticos para investigar as contribuições existentes sobre chatbots educacionais. Todos os estudos secundários que foram selecionados nesta pesquisa estão listados na Tabela 3. Essa tabela contém um identificador (ID) do artigo, assim como o local de publicação dos artigos. Conforme é possível observar na Tabela 3, os artigos foram publicados em diferentes veículos de divulgação científica, a maioria em periódicos. Somente o artigo *e1* [Paschoal et al. 2020] foi publicado em uma conferência. É possível que isso tenha alguma relação com o limite de páginas imposto pelas conferências da área e um número reduzido de páginas pode dificultar a discussão e a síntese necessária para apresentar esse tipo de estudo.

Os estudos secundários foram publicados entre os anos de 2019 e 2020, conforme pode ser constatado na Tabela 4. Cada um contemplou um determinado período de tempo.

<sup>6</sup>Mais informações disponíveis em: <https://parsif.al/>

**Tabela 3. Estudos secundários selecionados**

| ID do artigo | Título do estudo secundário   | Local de publicação                            |
|--------------|---|--|
| e1           | A Systematic identification of pedagogical conversational agents  | IEEE Frontiers in Education Conference         |
| e2           | A systematic literature review of intelligent tutoring systems with dialogue in natural language                              | IEEE Access                                    |
| e3           | Discussing with a computer to practice a foreign language: research synthesis and conceptual framework of dialogue-based CALL | Computer Assisted Language Learning            |
| e4           | How are you, chatbot? Evaluating chatbots in educational settings - Results of a literature review                            | Lecture Notes in Informatics                   |
| e5           | Rediscovering the use of chatbots in education: A systematic literature review  | Computer Applications in Engineering Education |

O artigo *e1* [Paschoal et al. 2020], por exemplo, colocou um critério nas buscas dos estudos primários para que somente estudos primários publicados entre 2007 e 2017 fossem analisados. Algo similar foi feito nos artigos *e2* [Paladines and Ramirez 2020] e *e5* [Pérez et al. 2020], com a justificativa de que as publicações mais recentes abordam chatbots baseados em tecnologias não obsoletas. Os demais estudos secundários não delimitaram um período de tempo para conduzir a análise. Ainda, na Tabela 4, a quantidade de estudos primários analisados em cada um dos estudos é indicada.

**Tabela 4. Período de publicação e análise dos estudos**

| ID do artigo | Ano de publicação | Período analisado | Número de estudos |
|--------------|-------------------|-------------------|-------------------|
| e1           | 2020              | 2007-2017         | 101               |
| e2           | 2020              | 1999-2017         | 49                |
| e3           | 2019              | 1982-2017         | 207               |
| e4           | 2019              | 2007-2018         | 25                |
| e5           | 2020              | 2005-2020         | 80                |

## QP2 – Quais são os escopos dos estudos secundários sobre chatbots educacionais?

A segunda questão de pesquisa teve a intenção de identificar o foco de cada estudo secundário. O artigo *e1* [Paschoal et al. 2020] retrata um mapeamento sistemático que analisou estudos primários que reportavam chatbots produzidos para serem adotados na educação formal. Estudos que não abordavam chatbots para serem utilizados em práticas feitas na educação formal foram deixados à parte do mapeamento com auxílio dos critérios de seleção definidos pelos autores. Em particular, o artigo explora informações sobre áreas e níveis educacionais em que os chatbots foram aplicados, além de descrever as estratégias utilizadas pelos chatbots para se comunicar com o usuário. Adicionalmente, algumas informações sobre aspectos de desenvolvimento (*i.e.*, técnica de projeto, linguagem de programação) e formas de representação dos chatbots foram estudadas.

O artigo *e2* [Paladines and Ramirez 2020] analisa características dos chatbots que são introduzidos em sistemas tutores inteligentes (STI). Neste sentido, o foco da análise restringe-se ao abordar o uso de chatbots em atividades educacionais oferecidas pelos sistemas tutores inteligentes, no qual o chatbots assume a responsabilidade de oferecer tutoriais para os estudantes por meio do diálogo textual ou falado. Em especial, diferentes aspectos sobre a integração dos chatbots em STI foram analisados, tais como: os tipos de STI que utilizam chatbots; as áreas do conhecimento em que há suporte de chatbots; os níveis de escolaridade em que os STI foram aplicados; as características pedagógicas dos

chatbots; as abordagens para NLU (Natural-Language Understanding) que são adotadas pelos chatbots; dentre outros.

Outra revisão sistemática que estreita a análise sobre estudos primários é apresentada em *e3* [Bibauw et al. 2019], na qual os chatbots educacionais são explorados em situações específicas de uso. Nesse caso, o estudo situa a sua análise no contexto de chatbots usados para instrução assistida por computador (*i.e.*, CALL - Computer-Assisted Language Learning). Os CALL permitem que o aluno pratique a conversação em uma língua estrangeira e os chatbots surgem para promover a prática da conversação sem restrições linguísticas, processando texto ou fala. Nessa revisão sistemática, os autores discutiram sobre o tipo de diálogo que é oferecido pelos chatbots, as características da interação, da instrução e da tecnologia usada por esse tipo de sistema de software.

O artigo *e4* [Hobert 2019] foca na avaliação de chatbots educacionais. Os autores desse estudo perceberam que há pesquisadores atuantes na pesquisa sobre chatbots educacionais com formação básica em diferentes áreas (*e.g.*, psicologia, computação, educação). Ao mesmo tempo, esses autores observaram que diferentes métodos para avaliação dos chatbots educacionais são utilizados pelos pesquisadores, em decorrência de suas áreas de formação. Desse modo, o foco da revisão sistemática foi em levantar os métodos usados para avaliar os chatbots educacionais. Em particular, a revisão direciona a análise para estudos primários que abordam chatbots educacionais que passaram por um processo de avaliação, identificando os objetivos das avaliações, os tipos de estudo realizados e os instrumentos que são usados em tais avaliações.

Por fim, com um objetivo mais amplo que os outros estudos selecionados, o artigo *e5* [Pérez et al. 2020] reporta uma revisão sobre chatbots na educação sem restringir a seleção para um domínio ou um tipo de uso do software. Ele busca discutir sobre como a tecnologia de chatbot está sendo implementada em ambientes educacionais, quais os serviços que os chatbots oferecem para esses ambientes, assim como quais as ferramentas e técnicas que podem ser adotadas na avaliação dos chatbots. Além disso, o estudo debate sobre situações de aprendizagem em que o chatbot pode apoiar o aprendizado do aluno.

### **QP3 – Quais questões de pesquisa foram investigadas nos estudos secundários sobre chatbots educacionais?**

Os estudos secundários delimitaram as suas questões de pesquisa, com base em seus escopos. A Tabela 5 reúne as questões de pesquisa investigada em cada estudo. Constatou-se que nenhuma questão de pesquisa se repete nos estudos, mas pode-se estabelecer relação com o conteúdo retratado pelas questões. Observa-se que há argumentações produzidas para compreender as perceptivas de uso do chatbot, quanto aos conhecimentos dos chatbots, os níveis escolares para os quais eles foram estabelecidos, bem como seu público-alvo.

Também é possível notar que os estudos secundários, em sua maioria, retratam questões sobre a tecnologia usada na construção dos chatbots educacionais. O artigo *e1* [Paschoal et al. 2020] contém questões de pesquisa sobre as técnicas de projetos e as linguagens de programação usadas no desenvolvimento de chatbots. Já o artigo *e2* [Paladines and Ramirez 2020] assumiu que os chatbots foram produzidos com abordagens de NLU e procurou abordar questões com a intenção de assimilar as abordagens de NLU que foram usadas para implementar os chatbots. Ainda, o artigo *e5* [Pérez et al. 2020] direcionou a atenção para a arquitetura e componentes gerais dos chatbots, abordando como eles são treinados para compreender a língua humana.

Outra questão discutida pelos estudos secundários é a avaliação dos chatbots educacionais. O artigo *e4* [Hobert 2019] abordou com exclusividade questões para compreender métodos para avaliar os chatbots que foram usados nos estudos primários, os objetivos das avaliações e aspectos que tem sido avaliados. Além dele, o artigo *e5* [Pérez et al. 2020] contemplou uma questão relacionada com a qualidade do chatbot e

técnicas de avaliação.

**Tabela 5. Questões de pesquisa dos estudos secundários**

| ID do artigo | Questões de pesquisa  |
|--------------|---|
| e1           | In which areas have conversation agents been investigated?<br>At which educational levels are conversational agents used?<br>Which mechanisms do conversational agents use to interact with students?<br>What design techniques are used in the development of pedagogical conversational agents?<br>What programming languages are used in the development of pedagogical conversational agents?<br>How have the conversational agents been graphically represented?   |
| e2           | What are the essential features of the existing ITSs that have natural language dialogue?<br>What is the main purpose of the tutoring dialogue in each system?<br>What are the instructional approaches and the support resources used by the ITSs with natural language dialogue?<br>What is the NLU approach implemented by each system?<br>How have the ITSs with natural language dialogue been evaluated empirically?  |
| e3           | What are the boundaries of the field of dialogue-based CALL, how can we define this field, and what have been its major traditions and evolutions?<br>How can we categorize and distinguish the different types of dialogue-based CALL systems that have been developed so far, from interactional, instructional and technological perspectives?<br>What types of research and empirical designs have been used to study the impact of dialogue-based CALL, and what insights have been gained on its effectiveness? |
| e4           | Which state-of-the-art approaches to evaluate pedagogical conversational agents are common?<br>Which implications can be derived for future evaluation studies of pedagogical conversational agents?  |
| e5           | What are the different types of educational and/or educational environment chatbots currently in use?<br>How do they affect student learning or service improvement?<br>What type of technology do they use and what learning result is obtained from each of them?<br>In what cases does a chatbot help learning under conditions similar to those of a human tutor?<br>Would it be possible to evaluate the quality of chatbots, and what techniques exist for that?  |

#### **QP4 – Quais os desafios e as oportunidades de pesquisa para a temática de chatbots educacionais?**

Os estudos secundários identificaram algumas lacunas, indicando desafios de pesquisa e algumas questões que carecem de esforços ou um maior entendimento na área. Na tentativa de sintetizá-las e oferecer esclarecimentos sobre as oportunidades de pesquisa, um levantamento dessas lacunas foi realizado. Para complementar esse levantamento, os autores envolvidos nesta pesquisa buscaram localizar algumas fragilidades nos resultados apresentados pelos estudos secundários e acrescentaram alguns questionamentos tendo como base as discussões feitas nos estudos secundários. Pôde-se organizar as oportunidades de pesquisa em três vertentes sobre: (i) uso dos chatbots na educação; (ii) aspectos técnicos dos chatbots; e (iii) qualidade da avaliação dos chatbots.

*Sobre o uso dos chatbots na educação:* há chatbots para diferentes níveis de escolaridade, mas faltam estudos que discutem o uso de chatbots na educação infantil

[Paschoal et al. 2020]. É possível que haja níveis de escolaridade mais adequados para uso dos chatbots, mas isso ainda não está claro na literatura. Também não existem estudos primários que investigaram os chatbots para alguns domínios específicos, como ensino de ciências agrárias [Paschoal et al. 2020]. Em contrapartida, há muitos esforços concentrados em desenvolver agentes para determinados assuntos. Por exemplo, diversos chatbots foram desenvolvidos para aprendizado de língua inglesa [Bibauw et al. 2019]. Inclusive, alguns estudos reproduzem esforços feitos anteriormente. Então, é preciso esclarecer porquê os autores repetem esforços já empregados em vez de reusar chatbots produzidos por seus pares. Os estudos precisam estimular o reuso de soluções estabelecidas. Talvez seja necessário melhorar as práticas de ciência aberta e encorajar os pesquisadores que desenvolvem chatbots a disponibilizar os códigos gerados no contexto de suas pesquisas.

*Sobre os aspectos técnicos dos chatbots:* Ao discutirem sobre o desenvolvimento dos chatbots, os estudos primários não esclarecem as técnicas de projeto [Paschoal et al. 2020] e, de modo geral, há uma disparidade nos detalhes sobre a construção do chatbot [Paladines and Ramirez 2020, Pérez et al. 2020]. Isso dificulta obter um panorama sobre as técnicas de projeto que têm sido usadas, os algoritmos e abordagens de NLU que são mais eficazes para gerenciar o diálogo falado e textual. Os estudos primários que descrevem sobre a implementação do chatbot, acabam indicando que existem muitas formas de implementação desse tipo de software e por meio dos estudos secundários não dá para perceber se existe alguma forma de desenvolvimento que seja mais apropriada de ser usada. Um entendimento sobre isso precisa ser desenvolvido, com o propósito de elucidar a forma de construção de chatbots capazes de manter conversas longas, credíveis, motivacionais e diversificadas.

*Sobre a qualidade da avaliação dos chatbots:* Alguns estudos procuram compreender o impacto dos chatbots na educação, mas ainda não está claro se há características presentes nesse tipo de software que se acentuam e influenciam no sucesso ou fracasso do uso do chatbot [Hobert 2019]. Os chatbots geralmente são avaliados com base na autopercepção do estudante e na usabilidade do sistema [Hobert 2019, Pérez et al. 2020]. Essas avaliações tendem a usar instrumentos não padronizados, o que pode ser tomado como um argumento para questionar a validade das avaliações [Wohlin et al. 2012]. Adicionalmente, não está claro se somente esses aspectos são suficientes para indicar que há valor pedagógico no uso do chatbot. Novos estudos precisam oferecer discussões sobre como avaliar o potencial do chatbot e, se necessário, apresentar formas alternativas para observar o valor do chatbot na educação. Inclusive, estudos experimentais que tem um propósito de observar o impacto do chatbot quando usado sob determinadas condições, são escassos [Bibauw et al. 2019]. Por serem raros e não apresentarem aspectos sistemáticos, dificultam as comparações.

É possível que algumas dessas lacunas sejam solucionadas por meio de novos estudos secundários criados para atender essas necessidades. Entende-se que existem algumas questões que não foram discutidas pelos estudos e carecem de contribuições. Por exemplo, o papel desempenhado pelo chatbot na atividade educacional ainda não está claro e se mantém desconhecido. Sabe-se que existem chatbots desenvolvidos para atuar como companheiro de aprendizagem, tutor e estudante, mas nenhuma taxonomia foi consolidada. Outro aspecto não investigado pelos estudos secundários remete às possibilidades afetivas dos chatbots. Acredita-se que existe espaço para compreender melhor como as modalidades de interação oferecidas pelos chatbots estão sendo exploradas (*i.e.*, comunicação verbal e não verbal).

#### **4. Limitações**

Alguns cuidados foram feitos ao longo do estudo, principalmente no que diz respeito ao processo de seleção dos estudos secundários. Nesse sentido, procurou-se descrever ao longo deste artigo todas as decisões tomadas. Alguns resultados obtidos ao longo do estudo terciário não foram descritos no artigo, em decorrência do limite de páginas.

Para complementar, um pacote de laboratório foi produzido e contém dados adicionais, como uma análise sobre os critérios de seleção usados pelos estudos secundários, com as palavras-chave que constituem as strings de busca usadas nos estudos secundários e as bases de dados que eles utilizaram. Além dessa análise complementar, a tabela de extração de dados e uma tabela com a análise realizada durante o processo de seleção dos estudos secundários foram incluídas. Um pacote de laboratório foi publicado no diretório Zenodo<sup>7</sup> e está disponível em: <https://doi.org/10.5281/zenodo.7093801>.

Apesar de algumas preocupações terem sido tomadas e documentadas, este estudo contém limitações. Uma delas tem relação com a identificação dos estudos secundários. Como o estudo foi realizado no último trimestre de 2021, alguns estudos secundários podem ter sido publicados e não foram identificados, porque não haviam sido indexados pelas bases de dados ou por terem sido publicados nesse período ou em um período posterior (*i.e.*, durante o primeiro semestre de 2022). Além dessa limitação, é importante ressaltar que embora os estudos terciários geralmente desenvolvam critérios de qualidade para analisar a qualidade dos estudos secundários, optou-se por não utilizar critérios de qualidade no presente estudo. Esses critérios facilitam classificar e selecionar estudos que conseguem atingir um certo nível de qualidade. Como esse estudo foi constituído por cinco estudos, optou-se por não utilizar critérios de qualidade. Todavia, em uma futura replicação deste estudo, espera-se que um conjunto de critérios de qualidade seja adotado.

## 5. Conclusões

Este artigo apresentou as contribuições que têm sido elaboradas na literatura sobre chatbots educacionais, dando atenção especial aos estudos secundários que agrupam evidências sobre o desenvolvimento e o uso desse tipo de software. Por meio dos estudos secundários, foi possível perceber que cada estudo foca em diferentes aspectos para sintetizar a literatura sobre chatbots educacionais, embora haja uma aproximação de alguns temas, como discussões sobre níveis de escolaridade e tópicos para os quais os chatbots têm sido desenvolvidos. Também, nota-se uma preocupação dos estudos secundários em tentar esclarecer como os chatbots são construídos e usados na educação. Dois estudos secundários acabaram retratando como os chatbots são explorados em contextos de uso específicos. Apesar disso, não existem esforços para compreender o uso do chatbot na educação a distância, um dos tópicos que tem ganhado espaço em fóruns de discussão da área [Han and Lee 2022]. Nos resultados deste estudo, algumas oportunidades de pesquisa foram propostas. Elas podem ser acolhidas pela comunidade para direcionar as pesquisas existentes. Ainda fica como sugestão para a comunidade que estudos secundários englobando a literatura cinza sejam produzidos, para suprir a ausência de um olhar para as soluções que têm surgido fora do contexto de pesquisas acadêmicas.

## Agradecimentos

Os autores gostariam de agradecer a CAPES - Código de Financiamento 001, ao CNPq (Processo nº 308445/2021-0) e a FAPESP (Processo nº 2020/09560-2) pelo apoio financeiro.

## Referências

- AbuShawar, B. and Atwell, E. (2015). Alice chatbot: Trials and outputs. *Computación y Sistemas*, 19(4):625–632.
- Allouch, M., Azaria, A., and Azoulay, R. (2021). Conversational agents: Goals, technologies, vision and challenges. *Sensors*, 21(24):8448.
- Ayedoun, E., Hayashi, Y., and Seta, K. (2015). A conversational agent to encourage willingness to communicate in the context of english as a foreign language. *Procedia Computer Science*, 60:1433–1442.

<sup>7</sup>Mais informações disponíveis em: <https://zenodo.org/>

- Benotti, L., Martínez, M. C., and Schapachnik, F. (2014). Engaging high school students using chatbots. In *Conference on Innovation and Technology in Computer Science Education*, page 63–68.
- Benotti, L., Martnez, M. C., and Schapachnik, F. (2018). A tool for introducing computer science with automatic formative assessment. *IEEE Transactions on Learning Technologies*, 11(2):179–192.
- Bezverhny, E., Dadteev, K., Barykin, L., Nemeshaev, S., and Klimov, V. (2020). Use of chat bots in learning management systems. *Procedia Computer Science*, 169:652–655.
- Bibauw, S., François, T., and Desmet, P. (2019). Discussing with a computer to practice a foreign language: research synthesis and conceptual framework of dialogue-based call. *Computer Assisted Language Learning*, 32(8):827–877.
- Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., and Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4):571–583.
- Frangoudes, F., Hadjjaros, M., Schiza, E. C., Matsangidou, M., Tsivitanidou, O., and Neokleous, K. (2021). An overview of the use of chatbots in medical and healthcare education. In Zaphiris, P. and Ioannou, A., editors, *Learning and Collaboration Technologies: Games and Virtual Environments for Learning*, pages 170–184, Cham. Springer International Publishing.
- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., and Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75:461–468.
- Han, S. and Lee, M. K. (2022). Faq chatbot and inclusive learning in massive open online courses. *Computers & Education*, 179:104395.
- Herpich, F., Nunes, F. B., Voss, G. B., and Medina, R. D. (2016). Three-dimensional virtual environment and npc: A perspective about intelligent agents ubiquitous. In *Handbook of Research on 3-D Virtual Environments and Hypermedia for Ubiquitous Learning*, pages 510–536. IGI Global.
- Hobert, S. (2019). How are you, chatbot? evaluating chatbots in educational settings – results of a literature review. *Lecture Notes in Informatics*, pages 259–270.
- Jeon, J. (2022). Exploring ai chatbot affordances in the efl classroom: young learners’ experiences and perspectives. *Computer Assisted Language Learning*, 0(0):1–26.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report 2.3, Keele University and University of Durham, Keele, UK.
- Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering – a systematic literature review. *Information and Software Technology*, 51(1):7–15.
- Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., and Linkman, S. (2010). Systematic literature reviews in software engineering – a tertiary study. *Information and Software Technology*, 52(8):792–805.
- Maicher, K., Danforth, D., Price, A., Zimmerman, L., Wilcox, B., Liston, B., Cronau, H., Belknap, L., Ledford, C., Way, D., et al. (2017). Developing a conversational virtual standardized patient to enable students to practice history-taking skills. *Simulation in Healthcare*, 12(2):124–131.
- Nguyen, H. D., Pham, V. T., Tran, D. A., and Le, T. T. (2019). Intelligent tutoring chatbot for solving mathematical problems in high-school. In *International Conference on Knowledge and Systems Engineering*, pages 1–6.

- Ong, E. C. J. and Soriano, Z. C. (2016). A conversational agent to shift students' affect state. In Baldoni, M., Baroglio, C., Bex, F., Grasso, F., Green, N., Namazi-Rad, M.-R., Numao, M., and Suarez, M. T., editors, *Principles and Practice of Multi-Agent Systems*, pages 86–97, Cham. Springer International Publishing.
- Paladines, J. and Ramirez, J. (2020). A systematic literature review of intelligent tutoring systems with dialogue in natural language. *IEEE Access*, 8:164246–164267.
- Paschoal, L. N., Loureiro Krassmann, A., Nunes, F. B., Morais de Oliveira, M., Bercht, M., Barbosa, E. F., and do Rocio Senger de Souza, S. (2020). A systematic identification of pedagogical conversational agents. In *IEEE Frontiers in Education Conference*, pages 1–9.
- Paschoal, L. N., Melo, S. M., Neves, V. d. O., Conte, T. U., and de SOUZA, S. d. R. S. (2022). An experimental study on a conversational agent in software testing lessons. *Informatics in Education*.
- Paschoal, L. N., Turci, L. F., Conte, T. U., and Souza, S. R. S. (2019). Towards a conversational agent to support the software testing education. In *Brazilian Symposium on Software Engineering*, page 57–66.
- Pereira, J., Fernández-Raga, M., Osuna-Acedo, S., Roura-Redondo, M., Almazán-López, O., and Buldón-Olalla, A. (2019). Promoting learners' voice productions using chatbots as a tool for improving the learning process in a mooc. *Technology, Knowledge and Learning*, 24(4):545–565.
- Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1 – 18.
- Pérez, J. Q., Daradoumis, T., and Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6):1549–1565.
- Ranoliya, B. R., Raghuvanshi, N., and Singh, S. (2017). Chatbot for university related faqs. In *International Conference on Advances in Computing, Communications and Informatics*, pages 1525–1530.
- Santos-Perez, M., Gonzalez-Parada, E., and Cano-garcia, J. M. (2013). Mobile embodied conversational agent for task specific applications. *IEEE Transactions on Consumer Electronics*, 59(3):610–614.
- Wik, P. and Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech communication*, 51(10):1024–1037.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.