



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE COMUNICAÇÃO
DEPARTAMENTO DE AUDIOVISUAIS E PUBLICIDADE

GISSELLE LAYLA ABDON DO VALE

**EXPLORANDO OS DESAFIOS DA COMUNICAÇÃO HUMANO-IA NA
EDUCAÇÃO: UMA REVISÃO SISTEMÁTICA DA LITERATURA**

BRASÍLIA

2023

GISSELLE LAYLA ABDON DO VALE

**EXPLORANDO OS DESAFIOS DA COMUNICAÇÃO HUMANO-IA NA
EDUCAÇÃO: UMA REVISÃO SISTEMÁTICA DA LITERATURA**

Trabalho de conclusão de curso apresentado à Faculdade de Comunicação (UnB), como requisito parcial à obtenção do título de Bacharel em Comunicação Social com habilitação em Publicidade e Propaganda.

Orientador: Priscila Monteiro Borges

Coorientador: Cleidson de Souza

BRASÍLIA

2023

GISSELLE LAYLA ABDON DO VALE

**EXPLORANDO OS DESAFIOS DA COMUNICAÇÃO HUMANO-IA NA
EDUCAÇÃO: UMA REVISÃO SISTEMÁTICA DA LITERATURA**

Trabalho de conclusão de curso apresentado à Faculdade de Comunicação (UnB), como requisito parcial à obtenção do título de Bacharel em Comunicação Social com habilitação em Publicidade e Propaganda.

Orientador: Priscila Monteiro Borges

Coorientador: Cleidson de Souza

BANCA EXAMINADORA

Profa. Dra. Priscila Monteiro Borges

Orientadora

Prof. Dr. Wladimir Ganzelevitch Gramacho

Examinador

Prof. Dr. Cleidson Ronald Botelho de Souza

Examinador

Profa. Dra. Carina Luisa Ochi Flexor

Suplente

AGRADECIMENTOS

Gostaria de agradecer a Deus, ao meu guia espiritual e a minha família espiritual. Agradeço pelos momentos felizes e tenho certeza de que sem eles não teria conseguido enfrentar os mais difíceis. Agradeço-lhes por permanecerem ao meu lado ao longo de todos esses anos. Esta monografia é uma etapa muito importante em minha vida, resultante de cinco anos repletos de altos e baixos e, em todos eles, quando faço uma prece, sinto-me abraçada.

À minha mãe. Minha melhor amiga, para sempre. Por seu apoio em todas as minhas decisões, pela generosidade, pelo exemplo de disciplina e determinação. Ao Jean que apareceu em nossas vidas para acrescentar e já fez toda a diferença. À tia Dinha e à minha avó Raimunda, que sempre se fizeram presentes, mesmo à distância.

Aos meus amigos Ryan e Maria Clara. Maria Clara tem uma forma única de ver o mundo, que faz com que eu também enxergue além. Obrigada por todas as metáforas que me fazem repensar minhas incertezas e preocupações. Ao Ryan, meu companheiro de graduação e meu amigo parceiro de momentos inesquecíveis. Obrigada por seu espírito acolhedor e por tornar a vida mais leve.

Aos meus orientadores, professora Priscila, minha prima Luciana e professor Cleidson, que me orientaram, acalmaram e incentivaram durante todo esse processo. Ao Leo e à Juliana pelas oportunidades de aprendizado, pelos conselhos valiosos e pela humildade, em que busco me espelhar para o resto da minha vida.

RESUMO

Vivemos uma era marcada por mudanças tecnológicas significativas e a Inteligência Artificial (IA) é um dos principais vetores dessa transformação. Seu impacto em diversos setores da sociedade já se mostram visíveis, especialmente no campo da educação. A interação entre humanos e IA em linguagem natural, com os avanços alcançados pelos grandes modelos de linguagem, suscita questões relevantes sobre a comunicação eficaz com máquinas. Dessa forma, este trabalho é uma revisão sistemática da literatura que se propõe a investigar e mapear os desafios comunicacionais na interação humano-IA, com foco na aplicação de grandes modelos de linguagem na educação. A revisão contemplou a análise de 28 artigos analisados com o método qualitativo, utilizando o método qualitativo de análise temática teórica com codificação aberta. O estudo identificou 8 temas principais (Acessibilidade; Relação Humano-IA; Literacia em IA; Engenharia de *Prompts*; Compreensão Contextual; Arquitetura, Interface e *Output*; Estilo Conversacional; e Questões Éticas), compostos por 31 subtemas, os quais foram minuciosamente discutidos e explorados.

Palavras-chave: *Large Language Models* (LLMs), Educação, Comunicação, Inteligência Artificial na Educação (AIEd), *chatbots*, agentes conversacionais, *Natural Language Processing* (NLP).

ABSTRACT

We live in an era defined by intense technological shifts, and Artificial Intelligence (AI) stands at the forefront of this transformation. Its impact on various sectors of society is already clear and particularly notable in the field of education. The interaction between humans and AI in natural language, with the leaps achieved by large language models, raises relevant questions about effective communication with machines. Thus, this work is a systematic literature review that aims to investigate and map the communicational challenges in human-AI interaction, focusing on the application of large language models in education. The review encompassed the analysis of 28 articles using a qualitative approach, employing a theoretical thematic analysis method with open coding. The study identified 8 main themes (Accessibility; Human-AI Relationship; AI Literacy; Prompt Engineering; Contextual Understanding; Architecture, Interface, and Output; Conversational Style; and Ethical Issues), composed of 31 subthemes, which were meticulously discussed and explored.

Keywords: Large Language Models (LLMs), Education, Communication, Artificial Intelligence in Education (AIEd), chatbots, conversational agents, Natural Language Processing (NLP).

LISTA DE FIGURAS

Figura 1 - Visão Geral dos Principais Campos da IA (DEEPLARNING.AI, 2019).....	14
Figura 2 - Etapas da Revisão Sistemática da Literatura.....	18
Figura 3 - Planejamento da Revisão.....	18
Figura 4 - Estratégia de Busca.....	20
Figura 5 - Condução da Revisão.....	21
Figura 6 - Fluxograma de Seleção de Artigos.....	24
Figura 7 - Mapa Temático dos Desafios Comunicacionais.....	25
Figura 8 - Modelo Dialógico de Comunicação.....	26
Figura 9 - Desafios Comunicacionais Relacionados ao Usuário.....	27
Figura 10 - Desafios Comunicacionais Relacionados ao LLM.....	38
Figura 11 - Desafios Comunicacionais Relacionados a Questões Éticas.....	46

LISTA DE TABELAS

Tabela 1 - Critérios de Inclusão e Exclusão.....	21
Tabela 2 - Retornos da Estratégia de Busca.....	23
Tabela 3 - Lista de Artigos Selecionados.....	68

SUMÁRIO

1. INTRODUÇÃO.....	9
2. ESTADO DA QUESTÃO.....	10
3. PROBLEMA DE PESQUISA.....	12
4. JUSTIFICATIVA.....	12
5. OBJETIVOS.....	12
5.1 Objetivo Geral.....	12
5.2 Objetivos Específicos.....	13
6. QUADRO REFERENCIAL TEÓRICO.....	13
6.1 Inteligência Artificial.....	13
6.2 Agentes Conversacionais Baseados Em LLMs.....	14
6.3 Inteligência Artificial na Educação (AIED).....	16
7. METODOLOGIA.....	17
7.1 Planejamento da Revisão.....	18
7.2 Condução da Revisão.....	21
7.3 Apresentação dos Resultados.....	23
8. RESULTADOS.....	25
8.1 Acessibilidade.....	27
8.1.1 Acesso aos Chatbots e Justiça Social.....	28
8.1.2 Habilidades Linguísticas.....	28
8.2 Relação Humano-IA.....	29
8.2.1 "Inteligência".....	29
8.2.2 Antropomorfização da IA e Substituição dos Humanos.....	30
8.2.3 Dependência Excessiva em Sistemas de IA.....	31
8.3 Literacia em IA.....	32
8.3.1 Dificuldades na Adoção de Novas Tecnologias.....	32
8.3.2 Alfabetização em IA.....	33
8.3.3 Treinamento de Professores e Escolhas Pedagógicas.....	33
8.3.4 Treinamento de Alunos e Desenvolvimento de Habilidades Holísticas.....	34
8.4 Engenharia de Prompts.....	35
8.4.1 Habilidades de Questionamento e Estratégias de Prompt.....	36
8.4.2 Limitações das Interfaces.....	37
8.5 Compreensão Contextual.....	38
8.5.1 Negasystem.....	39
8.5.2 Desassociação entre Objetos, Significados e suas Representações.....	39
8.5.3 Imitação do Discurso.....	40
8.6 Arquitetura, Interface e Output.....	40
8.6.1 Qualidade dos Dados de Entrada.....	41
8.6.2 Falta de Transparência e Explicabilidade.....	41
8.6.3 Acurácia do Output.....	41

8.6.4 Alucinações.....	43
8.6.5 Não-Determinismo.....	43
8.6.6 Calibração de Confiança.....	44
8.6.7 Problemas Técnicos.....	44
8.7 Estilo Conversacional.....	44
8.7.1 Tom da Linguagem.....	45
8.7.2 Nível de Proatividade.....	45
8.7.3 Neutralidade do Output.....	46
8.8 Questões Éticas.....	46
8.8.1 Geração de Conteúdo Tóxico e Enviesado.....	47
8.8.2 Reprodução de Visões Dominantes.....	47
8.8.3 Atribuição do Uso de Chatbots.....	48
8.8.4 Plágio e Violação de Direitos Autorais.....	48
8.8.5 Fabricação de Referências e Citações.....	49
8.8.6 Privacidade e Coleta de Dados.....	49
8.8.7 Políticas sobre a Utilização de IA Generativa na Educação.....	50
9. DISCUSSÃO.....	50
9.1 Prompting Large Language Models with The Socratic Method (CHANG, 2023).....	51
9.2 How General-Purpose is a Language Model? Usefulness and Safety with Human Prompters in the Wild (CASARES et al., 2022).....	51
9.3 Don't Just Tell Me, Ask Me: AI Systems That Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy Over Causal Ai Explanations (DANRY et al., 2022).....	53
9.4 Co-Writing with Opinionated Language Models Affects Users' Views (JAKESCH ET AL., 2023).....	54
9.5 "I Think You Might Like This": Exploring Effects of Confidence Signal Patterns on Trust in and Reliance on Conversational Recommender Systems (RADENSKY et al., 2023).....	55
9.6 Prompting Meaning: A Hermeneutic Approach to Optimising Prompt Engineering with Chatgpt (HENRICKSON, MEROÑO-PEÑUELA, 2023).....	56
10. CONCLUSÃO, LIMITAÇÕES E TRABALHOS FUTUROS.....	57
11. REFERÊNCIAS.....	59
12. APÊNDICE.....	68

1. INTRODUÇÃO

Estamos vivendo em uma época marcada por transformações tecnológicas sem precedentes e a Inteligência Artificial (IA), que já faz parte do nosso cotidiano, desempenha um papel central nesse cenário. O livro *The Big Nine* (WEBB, 2019, p. 2) aponta que a IA “é a tecnologia sobre a qual nosso futuro está sendo construído [...]”¹ e acrescenta, ainda, “[...] é a terceira era da computação. Estamos no meio de uma transformação significativa, assim como a geração que viveu a Revolução Industrial”.

O avanço do campo da IA não representa um único evento disruptivo, mas sim uma cascata contínua de disrupções e transformações, conforme exposto por Harari (2017). Essas mudanças têm implicações profundas no campo da educação, gerando a necessidade de um retreinamento radical ao longo da vida. A revolução educacional prevista por Harari desafia a composição do currículo das escolas e universidades, bem como afeta os profissionais já inseridos no mercado de trabalho.

Desse modo, surgem discussões acerca da ameaça aos empregos tradicionais devido à automação das funções. Adicionalmente, fala-se do surgimento constante de novas profissões que demandam qualificações adequadas, contando com novas habilidades e conhecimentos atualizados. Essas transformações generalizadas impulsionam a necessidade de uma educação adaptada à nova realidade, capaz de acompanhar as demandas do mercado de trabalho.

Diante desse panorama, o foco deste trabalho é investigar os desafios das interações entre agentes comunicacionais e usuários em contextos educacionais, sob a perspectiva do campo da comunicação. Serão examinados os diversos contextos de comunicação, ou seja, as interações humano-IA que ocorrem por meio da linguagem natural, com o objetivo de promover a aprendizagem. Por exemplo, sistemas de tutoria inteligente, agentes pedagógicos, *chatbots*, entre outros. Essa análise permitirá compreender como a IA pode ser incorporada de maneira eficaz nos processos educacionais, considerando as necessidades dos usuários e os desafios decorrentes dessas interações.

¹ Todas as traduções das citações de obras em língua estrangeira foram realizadas pela autora deste artigo, com o auxílio do modelo de linguagem ChatGPT (GPT-3.5), da OpenAI, e o Bard (Gemini Pro), da Google.

2. ESTADO DA QUESTÃO

Os esforços para desenvolver uma Inteligência Artificial Centrada no Ser Humano (HCAI, do inglês, *Human Centered Artificial Intelligence*) estão entre os principais desafios do campo da IA (OZMEN et al., 2023). Yang (2020) destaca a importância de traduzir os avanços no contexto dos laboratórios para a integração da IA na forma de sistemas sociotécnicos no mundo real, que realmente atendam às necessidades humanas.

Ao mesmo tempo, Følstad e Brandtzæg (2017) observam a transição das interfaces gráficas baseadas em cliques e rolagens para interfaces de linguagem natural. As conversas passam a ser o foco principal do design. Segundo os autores, isso representa uma transição do design de layouts visuais para o design conversacional, no qual a interação depende da compreensão das necessidades e da interpretação efetiva das entradas do usuário. Portanto, é necessário entender como adaptar dinamicamente a interface à medida que o diálogo com o usuário se desenvolve.

No livro *A Inteligência Artificial é Inteligente?*, Santaella (2023) dedica um capítulo à aprendizagem humana para posteriormente introduzir outro acerca da aprendizagem de máquina. Para ela, dadas as reflexões acerca das principais teorias da aprendizagem, “[...] a aprendizagem se constitui em um dos mais importantes fatores de transformação e crescimento do humano como ser social” (SANTAELLA, 2023, p. 125).

Além do valor intrínseco da aprendizagem, Kaufman (2022, p. 18) acrescenta que, no contexto da automação inteligente apoiada por tecnologias de IA, os investimentos em educação são “imperiosos para qualificar e requalificar os trabalhadores, estando eles ou não empregados”. A autora defende que a educação tem o potencial de evitar que uma parcela significativa da sociedade fique excluída dos benefícios resultantes do progresso tecnológico.

No que se refere às tecnologias no campo da educação, Roda, Angehrn e Nabeth (2001) afirmam que elas devem ser capazes de facilitar o acesso, acelerar e personalizar o processo de aprendizagem, bem como oferecer um ambiente de aprendizagem mais rico. Sendo assim, a IA pode ser uma aliada importante para a educação e sua aplicação nessa área tem sido estudada por cerca de 30 anos (ZAWACKI-RICHTER et al., 2019).

Essas tecnologias podem auxiliar tanto os alunos quanto os professores, levando em conta problemas como a falta de suporte individual, especialmente nas universidades, onde muitas vezes há um desequilíbrio entre o número de professores e alunos (CHUN HO et al., 2018; WINKLER et al., 2019 apud SCHLIMBACH et al., 2022).

O *software* educacional está sendo equipado com ferramentas que se comunicam diretamente com os alunos e adaptam-se aos seus estados cognitivos e emocionais. Griol e Callejas (2016) mencionam que entre os ganhos da utilização de diálogos em linguagem natural estão: a concentração dos recursos cognitivos na própria tarefa de aprendizado, em vez de gastá-los na compreensão e utilização da interface dos aplicativos; e a presença do contexto e dimensão social que podem aumentar a motivação dos aprendizes.

Exemplos de técnicas em tecnologias de IA voltadas à educação são a utilização do diálogo socrático, incentivando o questionamento e a discussão; a adoção de suportes metacognitivos; a utilização e frameworks narrativos; a moderação da colaboração em grupo; e a utilização de modelos de simulação social, incorporando, por exemplo, um especialista, tutor ou colega virtual (LUCKIN et al., 2016). É observado também na literatura o impacto positivo da tutoria no aprendizado dos alunos, relativo ao ensino em sala de aula, porém os pesquisadores ainda investigam os componentes conversacionais que possam explicar esses ganhos (GRAESSER, LI, FORSYTH, 2014).

A aplicação de agentes conversacionais, apoiados pelos avanços em *Machine Learning (ML)*, em contextos educacionais permite aos aprendizes aplicarem regras semelhantes às interações sociais humano-humano. Nesse sentido, têm sido destacadas como áreas de pesquisa essenciais a serem exploradas: a adaptação da estrutura da comunicação e a antropomorfização dos agentes conversacionais para nortear diretrizes de design que facilitem uma interação fácil, flexível e natural a depender do contexto, tarefa e necessidades de cada usuário (SCHMITT et al., 2022).

Há outras questões pertinentes a serem consideradas no contexto das interações entre humanos e agentes conversacionais. São elas: as variações em riqueza de mídia (CARMICHAEL, 2022) e a confiança em relação à privacidade dos dados pessoais e à veracidade das informações fornecidas (DAĞLI, 2018). Muitas vezes, esses agentes operam como "caixas-pretas", sem oferecer explicações claras sobre como chegam a determinados resultados (GUNNING, 2019).

Por fim, busca-se explorar as implicações de uma educação datificada. Conforme descrito por Santaella (2023), nossas atividades digitais geram dados de natureza afetiva, cognitiva e física, contribuindo para a construção contínua de perfis que a autora denomina de "nossos gêmeos digitais". À medida que a comunicação com os agentes conversacionais evolui e eles adquirem um conhecimento mais aprofundado sobre nós, surge a possibilidade de personalizar ainda mais as interações e a comunicação no contexto estudado.

3. PROBLEMA DE PESQUISA

Quais são os desafios comunicacionais da utilização de agentes conversacionais baseados em *Large Language Models* (LLMs), em contextos educacionais, descritos por pesquisadores em literatura científica especializada?

4. JUSTIFICATIVA

Constata-se a importância de estratégias para dar suporte à nova realidade da educação, adaptando o sistema educacional às necessidades deste século. Destarte, no contexto apresentado, cabe a utilização de tecnologias como a Inteligência Artificial, entre outras, para apoiar e facilitar o processo de aprendizagem.

Ganha destaque o design conversacional e os desafios comunicacionais gerados pelas possibilidades de interações em linguagem natural baseadas em aplicações da IA. Assim, esta pesquisa levanta questões cruciais sobre os desafios em estabelecer essa comunicação, a fim de complementar as teorias e metodologias de aprendizado, apoiando as práticas dos professores e promovendo ganhos na aprendizagem e, possivelmente, no engajamento dos estudantes.

5. OBJETIVOS

5.1 Objetivo Geral

Investigar os desafios comunicacionais da utilização de agentes conversacionais baseados em LLMs em contextos educacionais descritos por pesquisadores em literatura científica especializada.

5.2 Objetivos Específicos

1. Mapear e coletar bibliografia sobre as formas de utilização de agentes conversacionais baseados em LLMs na educação, com ênfase na comunicação humano-IA;
2. Definir as principais questões discutidas pelos pesquisadores em relação a comunicação humano-IA em agentes conversacionais baseados em LLMs na educação;
3. Analisar e classificar os desafios comunicacionais da implementação dos *chatbots* baseados em LLMs aplicados na educação.

6. QUADRO REFERENCIAL TEÓRICO

6.1 Inteligência Artificial

Após a Segunda Guerra Mundial, começou-se a trabalhar em máquinas inteligentes. Alan Turing, reconhecido como o "pai da ciência da computação", possivelmente foi o pioneiro nesse campo. O site da IBM traça um breve histórico apontando que o artigo influente de Turing (1950), intitulado "*Computing Machinery and Intelligence*", apresentou uma pergunta fundamental: "as máquinas podem pensar?" A partir dela, ele propôs o chamado "Teste de Turing", em que um interrogador humano tenta distinguir entre as respostas em texto de um computador e de um ser humano.

Não há consenso acerca da definição de IA, assim como ainda é um desafio precisar o que é, de fato, a própria inteligência. Para Webb (2020), a IA é um sistema que toma decisões autônomas, já a IBM considera que a IA é "um campo que combina ciência da computação e conjuntos de dados robustos para possibilitar a resolução de problemas"². McCarthy (2004), em seu artigo "*What is Artificial Intelligence?*" propõe outra definição:

É a ciência e a engenharia de criar máquinas inteligentes, especialmente programas de computador inteligentes. Está relacionada à tarefa similar de usar computadores para compreender a inteligência humana, mas a IA não precisa se limitar a métodos que sejam biologicamente observáveis. (MCCARTHY, 2004, p. 2)

Ao passo que exploramos a complexidade de nossos cérebros, os mecanismos da consciência e da inteligência, também projetamos capacitar as máquinas com essas mesmas habilidades.

² <https://www.ibm.com/topics/artificial-intelligence>

Pode-se dizer que a grande meta dos desenvolvedores de IA é criar programas de computador que sejam capazes de aprender, resolver problemas, planejar para o futuro e alcançar objetivos tal como a inteligência humana (MCCARTHY, 2004). Essa seria a chamada “*General AI*”, projeto que ainda está no campo das ideias e da ficção científica. Hoje, são utilizadas sobretudo aplicações da “*Narrow AI*”, modelos treinados para realizar uma tarefa específica. As tecnologias atuais de IA baseiam-se principalmente em modelos estatísticos que calculam a probabilidade de ocorrência de eventos com base em dados (KAUFMAN, 2022).

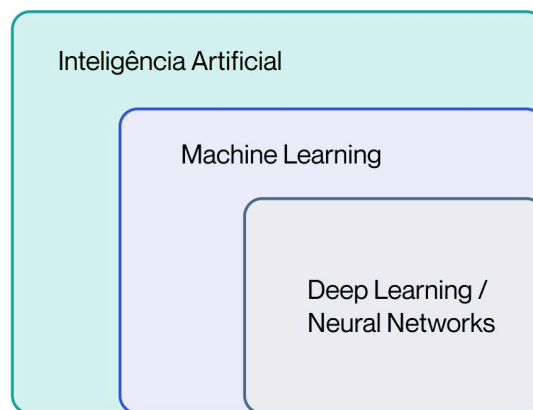


Figura 1 - Visão Geral dos Principais Campos da IA (DEEPLARNING.AI, 2019)

Ao adentrar o vasto campo da IA (Figura 1), encontra-se o subcampo denominado *Machine Learning* (ML). Samuel (2000) define ML como uma área de estudo que possibilita aos computadores aprenderem por meio da experiência, de maneira análoga aos seres humanos e animais, sem a necessidade de uma programação explícita. Dentro do ML, há um ramo conhecido como *Deep Learning* (DL), que se baseia em redes neurais.

De maneira sucinta, de acordo com o site da IBM, a diferença entre aprendizagem de máquina clássica (ou não profunda) e aprendizagem profunda reside na forma como os algoritmos aprendem. Na aprendizagem de máquina clássica, a máquina depende mais da intervenção humana para aprender e esse processo geralmente requer dados mais estruturados. Por outro lado, a DL automatiza grande parte do processo, sendo os algoritmos capazes de aprender diretamente de conjuntos de dados maiores e não estruturados, como texto e imagens.

6.2 Agentes Conversacionais Baseados Em LLMs

Com o aumento da sofisticação da IA e suas qualidades que parecem cada vez mais humanas, torna-se essencial estudar os desafios da interação humano-IA. Este trabalho concentrar-se-á

em uma das facetas desses desafios: a comunicação Humano-IA em agentes conversacionais, “programas de computador que interagem com usuários usando linguagem natural” (SHAWAR e ATWELL, 2007, p. 29), baseados em *Large Language Models* (LLMs).

A origem da IA conversacional remonta aos anos 1950, quando os modelos de linguagem eram bastante simples e as respostas da IA estavam restritas a regras predefinidas. O primeiro *chatbot*, ELIZA, data de 1966. ELIZA era um programa de computador de processamento de linguagem natural que simulava uma terapeuta, utilizando métodos que faziam com que os usuários tivessem a ilusão de que o programa os entendia (HE, 2023).

Para compreender o conceito de agentes conversacionais, é necessário primeiramente entender o que é *Natural Language Processing* (Processamento de Linguagem Natural). Em uma das principais obras didáticas no estudo de IA, Russel e Norvig (2010), apresentam dois motivos principais pelos quais almeja-se que agentes computacionais sejam capazes de processar linguagens naturais, como o inglês ou o português, que se distinguem de linguagens formais, como Java ou Python, utilizadas em programação. O primeiro motivo é a capacidade de comunicação com seres humanos, enquanto o segundo é a habilidade de adquirir informações a partir de linguagem escrita, considerando que existem mais de um trilhão de páginas recheadas de informações na web, sendo a grande maioria em linguagem natural.

Para o processamento da linguagem natural, uma abordagem utilizada é o emprego de modelos de linguagem, que têm a capacidade de prever a distribuição de probabilidade das expressões linguísticas. Os chamados *Large Language Models* (LLMs) se referem a modelos avançados que foram treinados em vastas quantidades de dados, na escala da internet (LEHMAN et al., 2022). Um exemplo proeminente de um LLM é o GPT-3, da OpenAI, contendo 175 bilhões de parâmetros (TAMKIN et al., 2021) e suas iterações, o GPT-3.5 e o GPT-4, que aumentaram significativamente quanto ao número de parâmetros e performance.

Ao interagir com os modelos de linguagem, as solicitações ou entradas fornecidas pelos usuários são denominadas ‘*prompts*’. Essas solicitações buscam gerar respostas específicas de acordo com a necessidade do usuário e, geralmente, consistem em perguntas, instruções ou afirmações que guiam a conversa. *Prompts* claros e concisos permitem a geração de respostas relevantes e úteis. Ao falar sobre *prompt engineering*, ou seja, engenharia de prompts, busca-se desenvolver e aprimorar o *prompt*, permitindo que o modelo produza resultados melhores. (DeepLearning.AI, Amazon Web Services, 2023)

Desde o *chatbot* ELIZA, os agentes conversacionais tiveram grandes avanços. Isso se deu principalmente por meio da utilização de grandes volumes de dados, combinados com melhorias significativas nos algoritmos de ML. Esses algoritmos fornecem diversas técnicas para a geração de textos semelhantes ao humano, como apontado por Hashana et al. (2023) em seu artigo sobre a aprendizagem profunda no ChatGPT. Apesar do avanço dessas tecnologias e do amplo leque de aplicações potenciais em várias áreas, seus desenvolvedores ainda enfrentam limitações e desafios como, por exemplo, o viés e a geração ocasional de respostas sem sentido, chamadas de "alucinações" (BAHRINI et al., 2023).

6.3 Inteligência Artificial na Educação (AIED)

Como uma das principais forças disruptivas deste século, a IA está sendo gradualmente incorporada ao currículo das escolas e é considerada não apenas como uma força motriz da quarta revolução industrial, mas também como um catalisador para a quarta revolução na educação (ZHAI et al., 2021). Nesse cenário, grandes empresas como Google, Amazon e Facebook estão investindo massivamente em produtos de Inteligência Artificial na educação (AIED). Este estudo se torna ainda mais relevante ao considerarmos o conceito de *Lifelong Learning*:

“Existe uma habilidade básica que está se tornando cada vez mais importante no universo tecnológico em constante mudança de hoje: ser capaz de aprender e se adaptar às novas habilidades e treinamentos que serão necessários”. (LAAL, 2013, p. 981)

O campo da AIED abrange uma ampla gama de teorias e metodologias provenientes de disciplinas como ciências cognitivas, educação, psicologia e comunicação, entre outras. A presença crescente desses sistemas inteligentes, adaptativos e personalizados nas escolas e universidades ao redor do mundo está sendo impulsionada pela análise de grandes quantidades de dados de estudantes, que servem como base para o desenvolvimento de ferramentas de apoio ao aprendizado e para a compreensão do próprio processo de aprendizagem.

Na literatura, constam diversas aplicações da AIED. Dentre elas estão presentes os sistemas de tutoria inteligentes, baseados em instruções ou em diálogo; suporte inteligente para a aprendizagem colaborativa; ambientes de aprendizado exploratórios, como realidade virtual em ambientes autênticos; avaliação automática da escrita; *chatbots*; e assistentes de ensino (HOLMES, BIALIK, FADEL, 2019; LUCKIN et al. 2016).

Ademais, o livro “*Artificial Intelligence in Education: Promise and Implications for Teaching and Learning*” (HOLMES, BIALIK, FADEL, 2019) categoriza as aplicações de AIEd em cinco dimensões: (a) tipo de alunos atendidos, (b) domínio de aprendizagem abrangido, (c) abordagens de aprendizagem facilitadas, (d) suporte de aprendizagem fornecido, e (e) suporte de ensino fornecido.

Uma característica relevante de múltiplos sistemas de AIEd é a presença de interações cada vez mais diretas e naturais entre os sistemas de IA (sejam eles *chatbots*, agentes pedagógicos, tecnologias de tutoria, entre outros) e os usuários.

7. METODOLOGIA

O objetivo desta pesquisa exploratória (GIL, 2013) é entender os desafios comunicacionais da interação com agentes conversacionais baseados em LLMs em contextos educacionais. Busca-se, desse modo, obter uma visão geral desse objeto de investigação para a formulação de problemas mais específicos e precisos em estudos posteriores. Quanto ao procedimento, esta é uma pesquisa bibliográfica, feita a partir do levantamento de referências teóricas já analisadas e publicadas por meios eletrônicos, como livros e artigos científicos. (FONSECA, 2002 apud MATOS e LERCHE, 2001, p. 31).

Neste trabalho, foi conduzida uma Revisão Sistemática da Literatura (RSL), um tipo de estudo secundário. Conforme descrito por Kitchenham e Charters (2007, p. VI), estudos secundários "revisam todos os estudos primários relacionados a uma questão de pesquisa específica, objetivando integrar e sintetizar as evidências relacionadas". Assim, visa-se identificar, avaliar e interpretar a pesquisa já existente (ibid.). No livro '*Introducing Systematic Reviews*', é mencionado também que "o objetivo da revisão sistemática é ter métodos explícitos, rigorosos e verificáveis" (GOUGH, OLIVER, THOMAS, 2017, p.6).

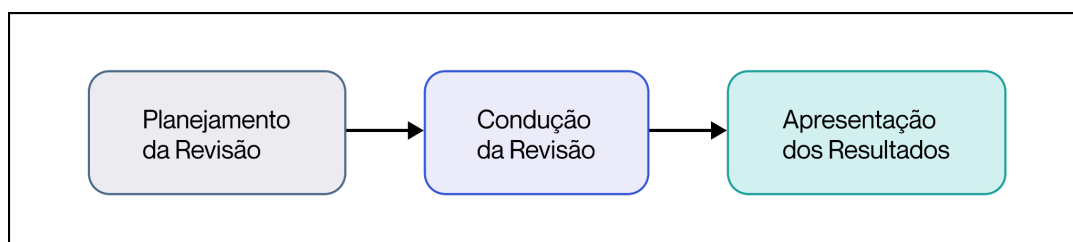


Figura 2 - Etapas da Revisão Sistemática da Literatura

Isto posto, para possibilitar que os leitores avaliem o rigor, a abrangência e a replicabilidade do processo, as estratégias de pesquisa deste trabalho foram documentadas detalhadamente. Assim, para realizar esta RSL, foram utilizadas como referência as diretrizes e o modelo de protocolo de revisão sistemática proposto por Kitchenham e Charters (2007). Em primeiro lugar, foi feito o planejamento, a condução da revisão e, em seguida, a apresentação dos resultados (etapas ilustradas pelas Figura 2).

7.1 Planejamento da Revisão

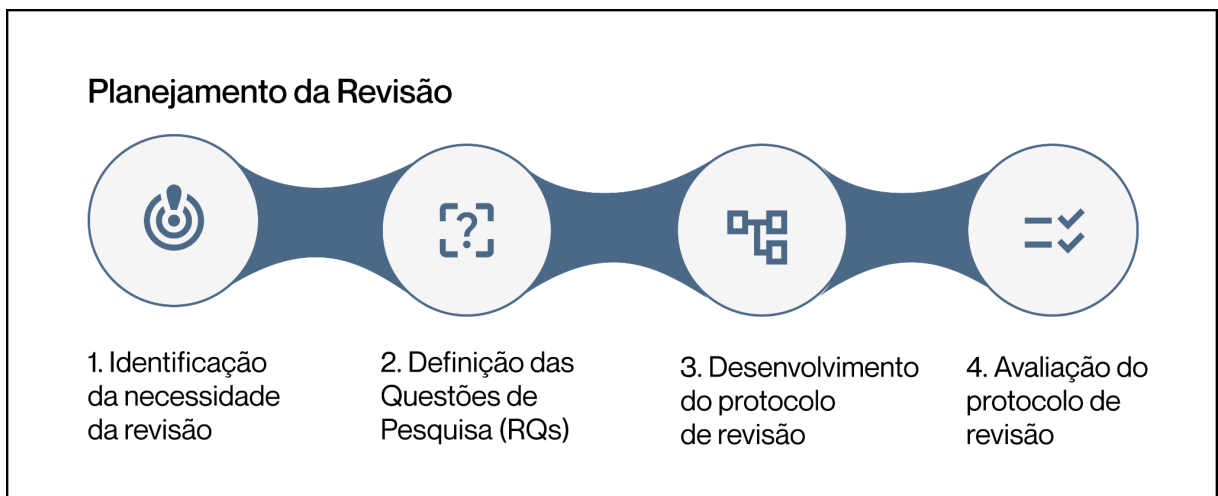


Figura 3 - Planejamento da Revisão

A primeira etapa do planejamento, conforme apresentado na Figura 3, foi a identificação da necessidade da revisão. Essa identificação foi baseada na definição do objeto de estudo de interesse e na realização de pesquisas preliminares na literatura, comumente chamadas de pesquisas de escopo (ver seções 1, 3 e 4).

A partir dessas pesquisas, a pergunta de pesquisa elaborada foi: *Quais são os desafios comunicacionais da utilização de agentes conversacionais baseados em Large Language Models (LLMs), em contextos educacionais?* (Etapa 2). Essa questão se situa na intersecção de três áreas: comunicação, educação e interação humano-computador (IHC). Além de contribuir para o posicionamento do campo da comunicação no contexto das tecnologias emergentes de IA e IA Generativa, também auxiliará a comunidade de educação e IHC no objetivo de projetar e avaliar *chatbots* educacionais.

Boland, Cherry e Dickson (2017, p. 158) afirmam que a pergunta de pesquisa é o destino da revisão e o protocolo de revisão é o mapa que orienta o pesquisador até esse destino,

definindo as atividades de pesquisa. Os autores enfatizam que esses dois componentes são essenciais para uma revisão bem-sucedida.

Neste estudo, o protocolo de revisão foi elaborado para especificar os métodos utilizados (etapa 3), assim como reduzir o viés de pesquisador, contando com os seguintes passos: (1) Determinação de estratégia para busca de estudos primários, incluindo termos de pesquisa e bases de dados; (2) Estabelecimento de critérios de seleção dos estudos primários, sendo estes definidos como pesquisas empíricas que investigam perguntas de pesquisa específicas; (3) Execução de procedimentos de seleção dos estudos primários; (4) Elaboração e aplicação de estratégia de extração de dados; (5) Análise e síntese dos dados extraídos; e (6) Confecção e formatação desta monografia como relatório final da revisão sistemática (Kitchenham e Charters, 2007).

O primeiro passo do protocolo de revisão, assim como o segundo e terceiro, foram orientados pela comparação com os métodos utilizados em outras doze RSLs no mesmo tema de estudo (CHIU et al., 2023; DEMERVAL et al., 2017; HOBERT, MEYER, 2019; HWANG, CHANG, 2021; KUHAIL et al., 2022; MARTHA, SANTOSO, 2019; OKONKWO, ADE-IBIJOLA, 2021; PÉREZ, DARADOUMIS, PUIG, 2020; PÉREZ-MARIN, 2021; RAMANDANIS, XINOGALOS, 2023; WINKLER, SOELLNER, 2021; WOLLNY et al., 2021). Para garantir a abrangência e a relevância da revisão, durante o primeiro passo, foram selecionadas como fontes de pesquisa as bases de dados eletrônicas mais utilizadas nesses artigos, a saber: IEEEExplore (utilizada por 7 estudos), ACM Digital Library (utilizada por 6 estudos), Scopus (utilizada por 7 estudos), SpringerLink (utilizada por 5 estudos) e Web of Science (utilizada por 5 estudos). Posteriormente, foi excluída a base de dados ScienceDirect, visto que não foi possível utilizar o mesmo *string* de busca utilizado nas demais, dadas as limitações de operadores booleanos estabelecidas pela base.

Os termos de pesquisa foram determinados a partir da pergunta de pesquisa, que foi dividida em quatro categorias de interesse: *chatbots*, *LLMs*, *Comunicação* e *Educação*. Para cada categoria, foram estabelecidas as palavras-chave mais importantes, bem como uma lista de alternativas, sinônimos e abreviações. Esses termos foram, então, refinados por meio de testes iniciais de busca até a construção da estratégia de busca final, ou seja, a sintaxe de pesquisa.

Como apontado pela figura 4, em cada base de dados, foram aplicadas duas estratégias de busca. Na primeira, foram combinados os termos das categorias *chatbots* e *LLMs* com os

termos da categoria *Educação*; e, na segunda estratégia, são combinadas as categorias *chatbots*, *LLMs* e *Comunicação*.

Para finalizar o planejamento da revisão, foi conduzida uma avaliação do protocolo. Esse último passo visou verificar a coerência interna do protocolo, ou seja, incluindo a congruência do *string* de busca em relação à pergunta de pesquisa, assim como a adequação dos dados a serem extraídos e do método de análise para responder efetivamente à questão proposta.

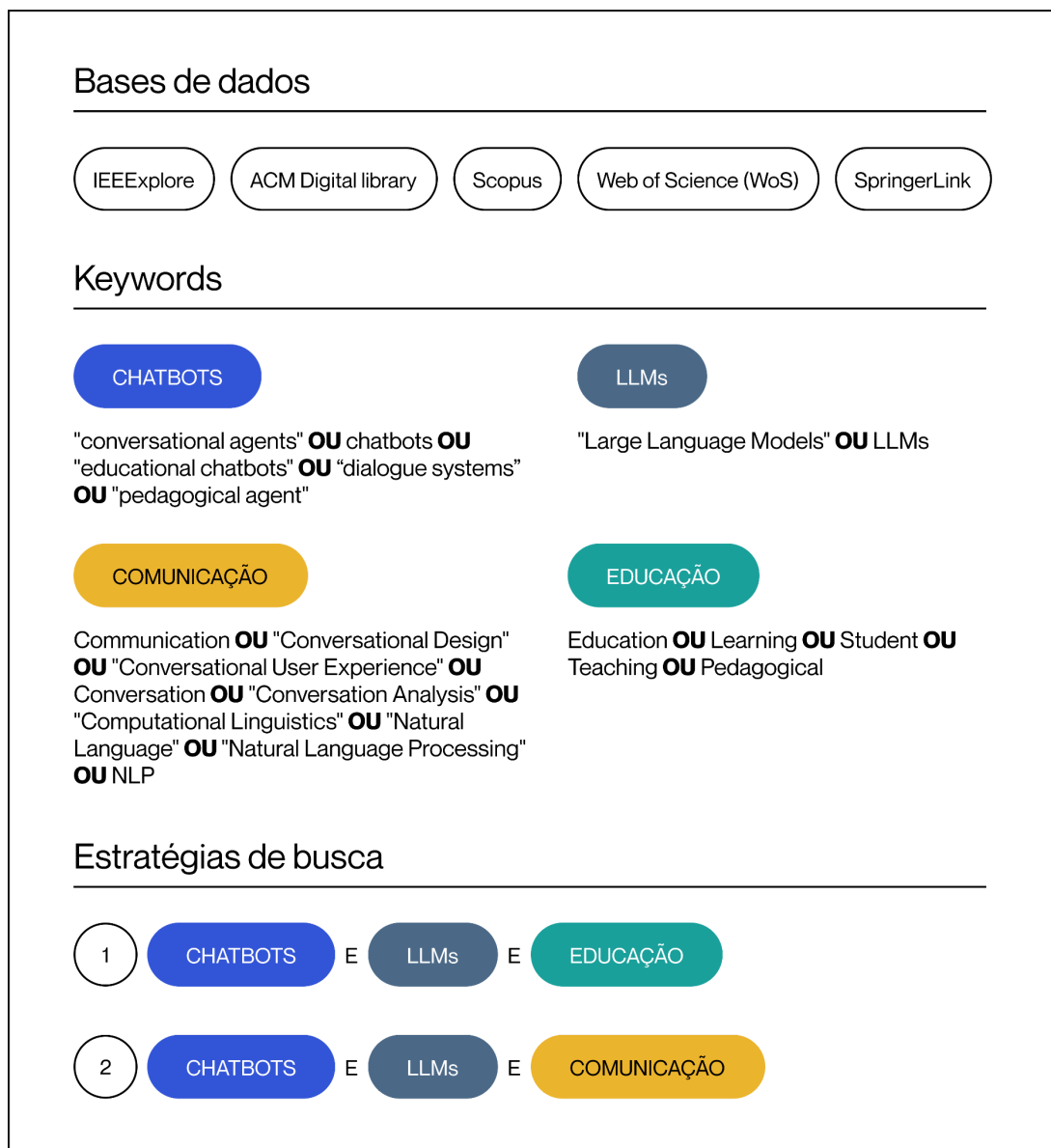


Figura 4 - Estratégia de Busca

7.2 Condução da Revisão

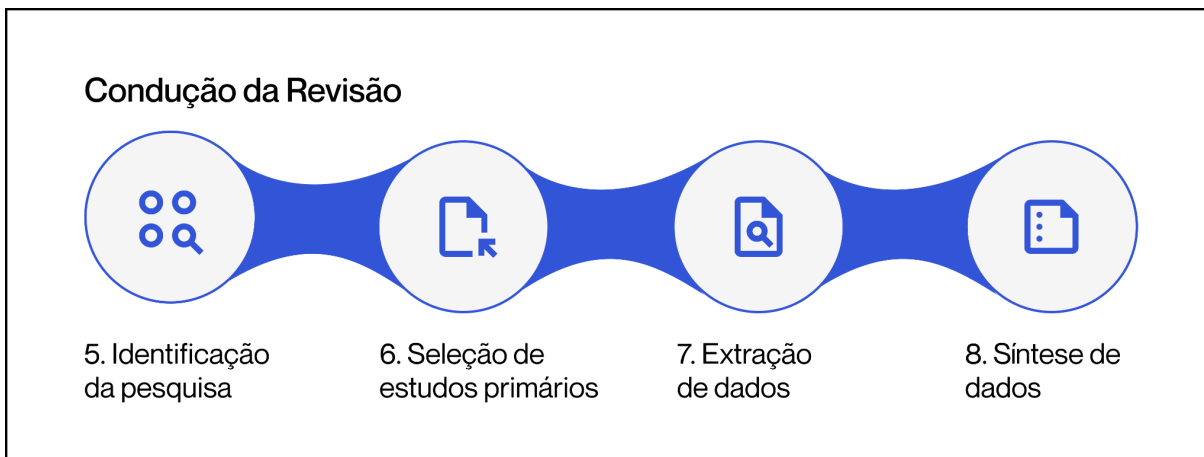


Figura 5 - Condução da Revisão

Após o desenvolvimento do protocolo, a condução da revisão propriamente dita foi iniciada. A pesquisa foi identificada (etapa 4) por meio da execução das estratégias de pesquisa e foram estabelecidos critérios de inclusão e exclusão (etapa 5) para avaliar cada estudo primário potencial e determinar quais seriam incluídos ou excluídos da RSL. Esses critérios, indicados na tabela 1, serão detalhados a seguir.

Tabela 1 - Critérios de Inclusão e Exclusão

Critérios de Inclusão	Critérios de Exclusão
CI1) Acesso	CE1) Estudos duplicados
CI2) Estudos primários e empíricos completos, de conferências ou revistas, revisados por pares	CE2) Relevância (I) Estudos exclusivamente teóricos, apresentação de nova ferramenta ou limitados a melhorar tecnicamente os <i>chatbots</i>
CI3) Data de publicação: entre junho de 2017 (publicação do artigo "Attention Is All You Need") e setembro de 2023	CE3) Relevância (II) Estudos em domínios ou contextos altamente específicos (Ex.: medicina)
CI4) Idioma: Inglês	CE4) Relevância (III) Artigos genéricos ou tangenciais (fora do escopo da Educação e Comunicação ou da pergunta de pesquisa)
	CE5) Percepção: Pesquisas dedicadas exclusivamente à descrição de como diferentes grupos reagiram à interação com a IA.

Em primeiro lugar, foram definidos os critérios de inclusão relacionados a questões práticas, como a possibilidade de acesso, o idioma e a data de publicação. O idioma inglês foi escolhido com base no maior número de pesquisas disponíveis, especialmente na área de tecnologia; e a data de publicação foi delimitada a partir da publicação do artigo 'Attention Is All You Need' (Vaswani et al., 2017), um marco significativo na área de Processamento de Linguagem Natural (NLP), considerando a introdução da arquitetura *Transformer*.

Já a preferência por artigos completos de conferências ou revistas baseia-se na qualidade da revisão por pares, atualidade e abrangência. Esses artigos oferecem maior confiabilidade, profundidade analítica e refletem o estado atual da pesquisa. E, por fim, foi aplicado o critério de relevância referente à abordagem do tema principal da pesquisa.

Entre os critérios de exclusão, encontram-se estudos duplicados, considerando a sobreposição das estratégias de pesquisa, assim como das bases de dados. Foram também excluídos os estudos não revisados por pares e a literatura cinzenta, que podem carecer de detalhamento metodológico, não garantindo a mesma qualidade e relevância desejadas.

Finalmente, foram estabelecidos os critérios de exclusão voltados à relevância dos artigos para o escopo da pesquisa. Sendo assim, foram excluídos os estudos exclusivamente teóricos, estudos em domínios altamente específicos, bem como artigos fora do escopo da educação, da comunicação ou da pergunta de pesquisa deste trabalho. O último critério de exclusão foi aplicado para descartar artigos cujas abordagens metodológicas se concentravam exclusivamente na coleta de percepções genéricas dos usuários durante suas interações com a IA.

A tabela 2 exibe a quantidade de artigos encontrados para cada estratégia de pesquisa nas diferentes bases de dados eletrônicas selecionadas. O total de artigos recuperados em todas as bases foi de 2173 artigos.

Já a figura 6 contém o fluxograma detalhado do processo de seleção de artigos. Este fluxograma abrange a identificação inicial dos artigos, assim como a triagem, seguindo os critérios de inclusão e exclusão estabelecidos previamente. Ao final deste processo, chegou-se a um total de 6 artigos relacionados ao aspecto comunicacional dos *chatbots* e 22 artigos referentes à implementação das LLMs no campo da educação, os quais foram incluídos na Revisão Sistemática da Literatura (RSL).

Tabela 2 - Retornos da Estratégia de Busca

Bases de dados	Estratégia 1	Estratégia 2	Total de Retornos
IEEEExplore	8	17	25
ACM Digital Library	135	139	274
SpringerLink	340	324	664
Scopus	622	562	1184
Web of Science (WoS)	15	11	26
Total	1120	1053	2173

Após a triagem dos artigos pelos critérios de inclusão e exclusão, os 28 artigos selecionados foram lidos e analisados de maneira detalhada. A etapa 6 envolveu a extração de dados. As informações foram minuciosamente extraídas de cada estudo primário através de uma abordagem de análise qualitativa conhecida como análise temática teórica com codificação aberta, seguindo a proposta de Braun e Clarke (2006). Essa metodologia foi selecionada com o objetivo de identificar, analisar e comunicar os padrões recorrentes, ou temas, observados nos dados analisados. A análise teórica foi escolhida tendo em vista a codificação para uma pergunta de pesquisa específica em relação aos dados coletados. Sendo assim, essa forma de análise tende a ser conduzida pelo interesse teórico ou analítico do pesquisador na área.

Nesse sentido, o objetivo foi identificar padrões e temas pertinentes aos desafios comunicacionais encontrados na interação com *chatbots* baseados em LLMs em contextos educacionais. A intenção foi, portanto, identificar conceitos, ideias ou padrões emergentes nos dados, fornecendo uma análise detalhada de uma questão específica presente nos dados (Braun e Clarke, 2006). Os resultados dessa análise serão apresentados na sessão seguinte.

7.3 Apresentação dos Resultados

A última etapa foi a apresentação dos resultados por meio da confecção e formatação desta monografia, seguida pela avaliação de uma banca examinadora. O trabalho foi formatado e os resultados da extração e síntese de dados estão descritos na seção seguinte.

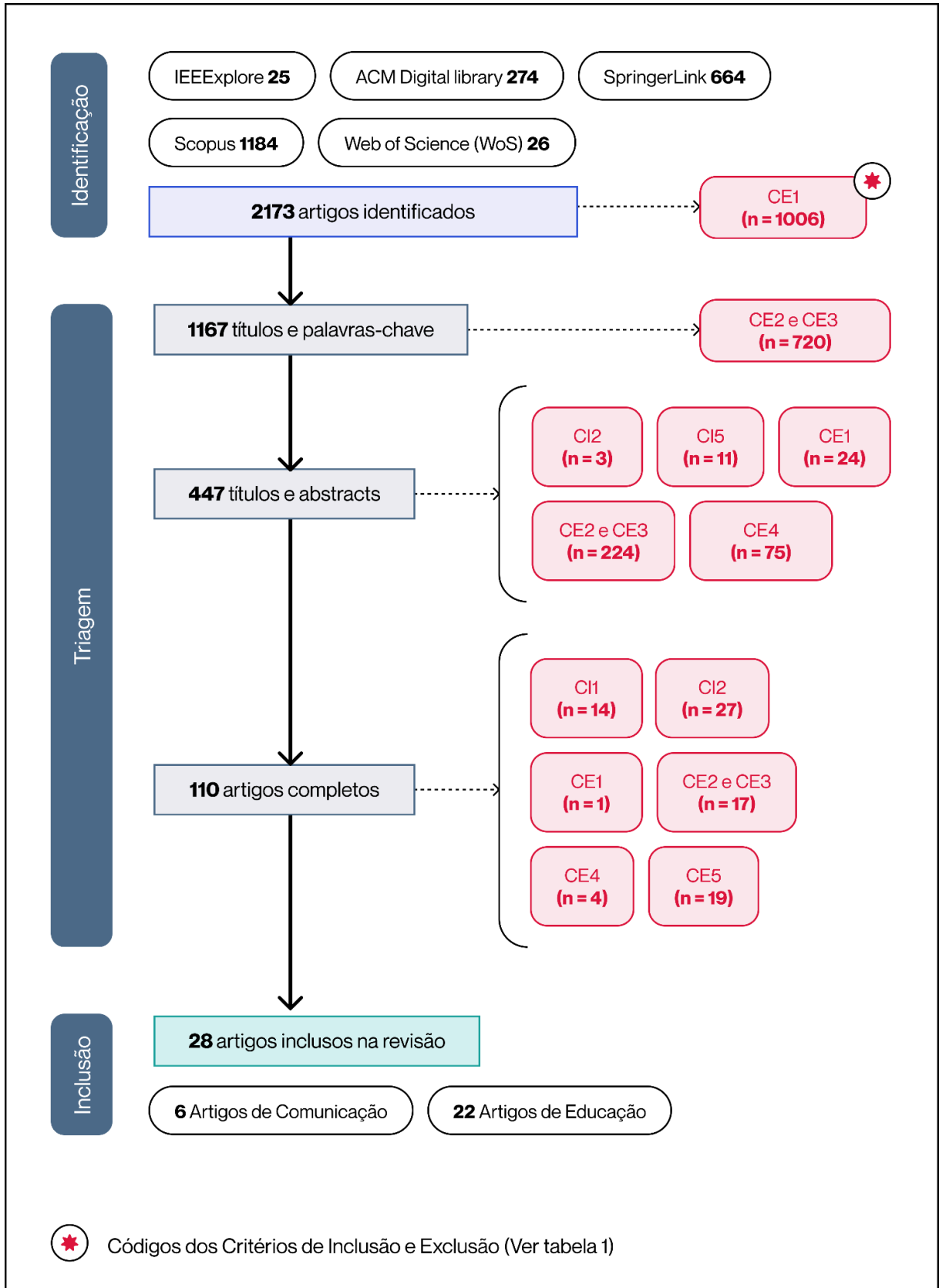


Figura 6 - Fluxograma de Seleção de Artigos

Nota de esclarecimento: neste estudo, foram empregados os modelos ChatGPT (GPT-3.5) e o Bard (Gemini Pro) como ferramentas de tradução e edição de texto, incluindo, por exemplo, a reformulação de frases para melhorar a fluidez, o estilo e a escolha das palavras. Todas as sugestões fornecidas pelos chatbots foram cuidadosamente analisadas e avaliadas criticamente antes de serem incorporadas, a fim de assegurar a consistência e a precisão das informações apresentadas.

8. RESULTADOS

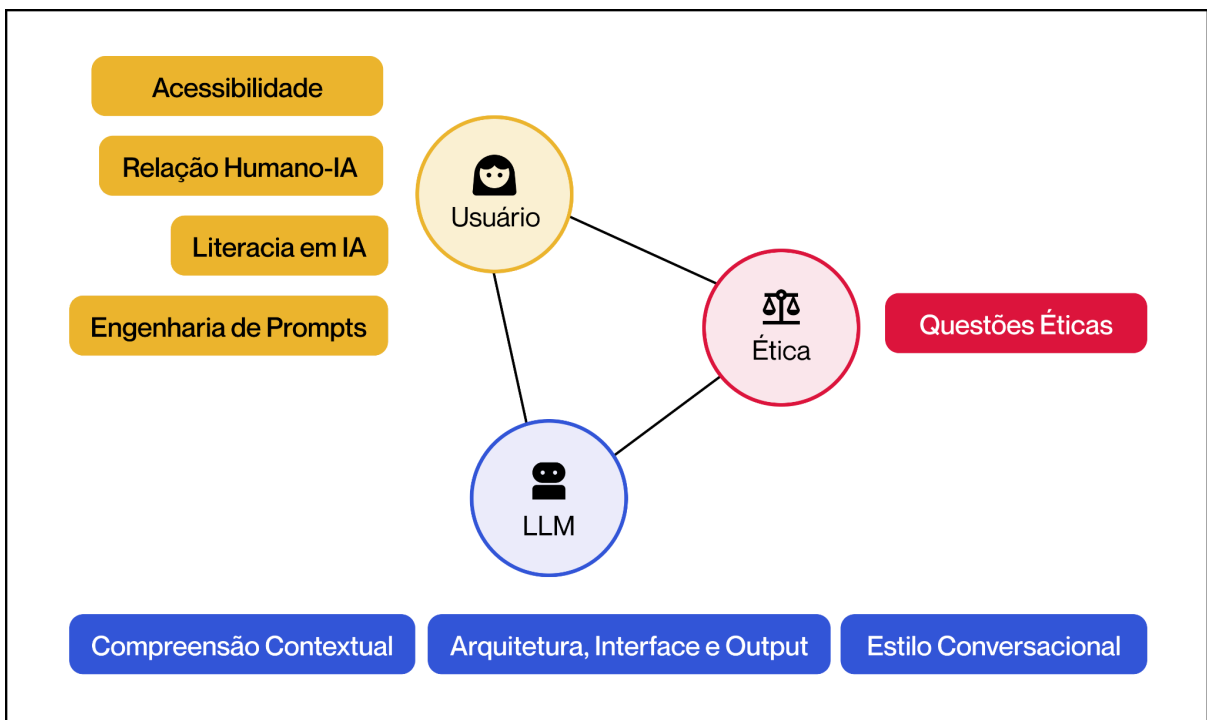


Figura 7 - Mapa Temático dos Desafios Comunicacionais

A partir da pergunta de pesquisa e a identificação dos temas, a investigação deu origem a um mapa temático dos desafios comunicacionais enfrentados (Figura 7). Assim, os desafios foram agrupados em oito temas mais abrangentes e seus subtemas, sendo cada um deles um desafio comunicacional distinto. Esses desafios foram nomeados e agrupados, garantindo a organização dos dados em grupos coerentes e evitando a sobreposição ou complexidade excessiva.

A partir do estabelecimento e do refinamento dos temas, buscou-se definir a essência de cada um e sua relação com a questão de pesquisa, bem como o conjunto de temas como um todo. Em seguida, será apresentada uma narrativa analítica acerca dos desafios comunicacionais

encontrados na interação com *chatbots* baseados em LLMs na educação. Com isso, visa-se descrever os dados, a hierarquia entre eles e relacioná-los à pergunta de pesquisa.

Para situar os diferentes desafios comunicacionais, foi utilizada uma adaptação do modelo comunicacional interativo (Figura 8).

O modelo interativo ou de interação da comunicação [...] descreve a comunicação como um processo no qual os participantes alternam posições como emissor e receptor, gerando significado ao enviar mensagens e receber feedback dentro de contextos físicos e psicológicos (SCHRAMM, 1997 apud ASHMAN, 2018)

Esse modelo contrasta com o linear, que retrata a comunicação como um processo unidirecional, onde um emissor transmite mensagens para um receptor (ASHMAN, 2018). Apesar de não abranger aspectos físicos e psicológicos, é possível estabelecer uma analogia entre o modelo interativo e a comunicação com máquinas. Contudo, vale notar que essa forma de comunicação pode estar situada em um ponto intermediário entre o modelo interativo e o modelo linear. Essa hipótese pode ser sustentada pela tendência da mensagem da IA ser consideravelmente influenciada pelo *prompt* do emissor, resultando em uma troca que não é completamente simétrica ou estritamente bidirecional, como no modelo interativo tradicional.

Isto posto, os desafios de comunicação identificados são ilustrados nas figuras 8, 9 e 10, agrupados de acordo com sua relação predominantemente com o usuário (Figura 8), com o LLM (Figura 9) e com questões éticas (Figura 10).

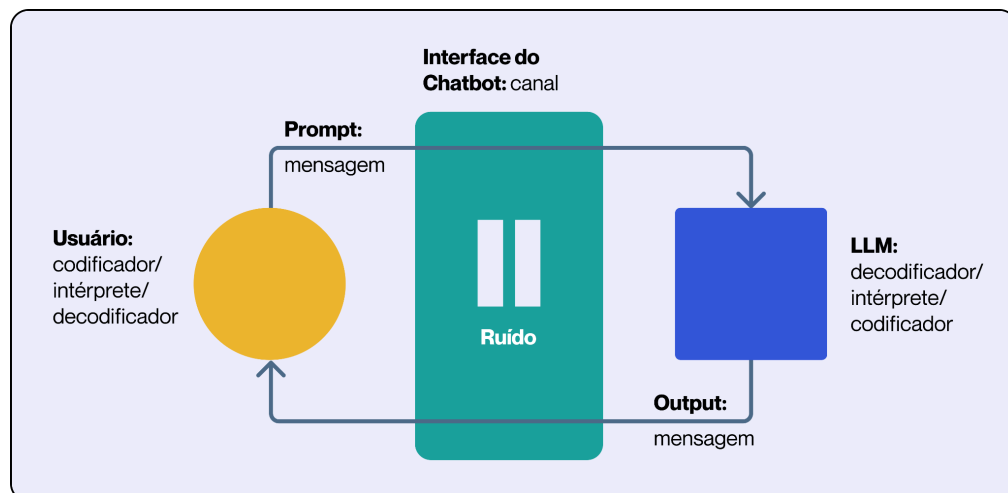


Figura 8 - Modelo Dialógico de Comunicação



Figura 9 - Desafios Comunicacionais Relacionados ao Usuário

8.1 Acessibilidade

O primeiro tema delineado foi "Acessibilidade", abordando a inclusão dos usuários, tanto no acesso aos *chatbots* quanto na facilidade de interação com eles. Nesse contexto, são considerados princípios de equidade de acesso e facilidade de interação com os agentes conversacionais.

Essa questão emerge como um desafio comunicacional, principalmente devido à disparidade de acesso a essa ferramenta entre distintos grupos de usuários, que pode acentuar questões relacionadas à justiça social. Adicionalmente, a diversidade de competências linguísticas e a

falta de suporte em determinados idiomas também exercem impacto significativo na eficácia da comunicação com esses agentes.

A ausência de acessibilidade pode acarretar disparidades no processo de aprendizado, restringindo a participação e o acesso equitativo ao conhecimento. Desse modo, tal cenário pode reforçar as desigualdades sociais e educacionais.

8.1.1 Acesso aos Chatbots e Justiça Social

A primeira subcategoria deste tema se refere ao acesso aos *chatbots* e à justiça social, ou seja, desafios mais abrangentes, como a falta de permissão ou proibição do acesso, assim como o banimento das ferramentas de IA. Luo et al. (2023) citam que a China tem o acesso bloqueado ao ChatGPT e que algumas universidades, regiões e países optaram por banir o uso da ferramenta em determinados contextos. Outro exemplo foi a reação rápida de proibição de uso do ChatGPT por parte das escolas unificadas de Nova York e Los Angeles para precaver-se ao risco de sua utilização para trapacear em tarefas (TLILI et al., 2023).

O artigo aponta que essas decisões podem criar uma divisão entre dois grupos e levantar preocupações sobre justiça social: "Aqueles que podem usar o ChatGPT, considerados os 'bem-sucedidos', enquanto aqueles que não podem são vistos como 'perdedores' (LUO et al., 2023, p. 8-9). Dessa maneira, o grupo de 'bem-sucedidos' "com habilidades para o usar o ChatGPT de forma eficaz e evitar suas armadilhas obteriam vantagens competitivas, potencialmente aprendendo melhor e mais rápido" (LUO et al., 2023, p. 12).

Além disso, Chaudhry et al. (2023) destacam as vantagens imediatas para os usuários de *bots* de Inteligência Artificial, uma vez que, na ausência de mecanismos de verificação de integridade acadêmica, essas ferramentas conferem uma vantagem injusta sobre aqueles que não as utilizam.

8.1.2 Habilidades Linguísticas

Observa-se as dificuldades de adaptação da ferramenta a diferentes níveis de alfabetização. Como o LLM requer um *prompt* preciso, até mesmo os adultos podem ter dificuldade em se comunicar com o ChatGPT. Para crianças na educação infantil e no ensino fundamental, cujas habilidades linguísticas e conhecimento contextual ainda estão em desenvolvimento, torna-se

ainda mais difícil estabelecer interações de alta qualidade (LUO et al., 2023; MURGIA et al., 2023).

"[...] É evidente que o ChatGPT precisa fornecer respostas mais adequadas para atender comunidades específicas de usuários, já que a legibilidade das respostas examinadas em nosso estudo não estava no nível considerado ideal pelas crianças envolvidas" (MURGIA et al., 2023, p. 314).

Outro ponto abordado refere-se aos *chatbots* não possuírem as mesmas habilidades em todos os idiomas, podendo ocasionar falta de estrutura e erros gramaticais. Exemplos disso são os idiomas com baixos recursos ou mais distantes, tais como o chinês e o romeno (RUDOLPH, TAN, TAN, 2023a). Observa-se também que, ao operar em português, o Bing Chat "dependia de fontes de internet de qualidade inferior, afetando adversamente a qualidade de suas respostas" (VASCONCELOS, SANTOS, 2023, p. 13).

8.2 Relação Humano-IA

Essa categoria aborda as crenças e atitudes dos usuários em relação à IA, examinando como essas percepções influenciam a interação e comunicação com esses sistemas. Considera-se que a maneira como os usuários percebem a IA pode impactar diretamente a confiança depositada nos *chatbots*, em sua utilidade e na forma como eles recebem e interpretam as mensagens geradas por esses sistemas (os *outputs*). Vale ressaltar que a percepção distorcida ou equivocada sobre as capacidades e limitações da IA pode suscitar o uso inadequado ou subutilização desses recursos tecnológicos, sobretudo no contexto educacional.

Além disso, as dinâmicas e interações presentes nos ambientes educacionais exigem uma integração sensível e cuidadosa dos *chatbots*, visando equilibrar o uso da tecnologia e preservar as habilidades e interações humanas fundamentais. Nesse sentido, os desafios apontados são: a possível dependência excessiva em sistemas de IA e a substituição do papel humano.

8.2.1 "Inteligência"

A definição do termo "inteligência" tem sido objeto de debate entre especialistas em IA, e segundo a visão de LUO et al. (2023), seria mais apropriado denominá-la como um recurso para o "Aumento de Inteligência" humana, em vez de considerá-la como inteligente ou como

substituta das ações realizadas por educadores e cuidadores. Luo et al. (2023) acrescentam que a IA só pode agir de acordo com objetivos comandados definidos por humanos, que podem ser interrompidos a qualquer momento de acordo com a decisão humana. Essa perspectiva posiciona a IA como uma ferramenta complementar, sob a supervisão e controle dos humanos.

Frick (2023) também argumenta contra a utilização da nomenclatura "artificialmente inteligentes" para esse gênero de sistema. De acordo com o autor, é necessário posicionar a IA como uma ferramenta, não como criatura. Sendo assim, a "[...] forma mais precisa de entender o que estamos construindo hoje é como uma 'forma inovadora de colaboração social'" (LANIER, 2023 apud FRICK, 2023, p. 9).

8.2.2 Antropomorfização da IA e Substituição dos Humanos

Esse código expande a discussão para além da dimensão da inteligência, que é geralmente o foco principal das comparações entre a IA e os seres humanos. Apesar de o ChatGPT-4 ser multimodal e interagir por meio da linguagem natural, ele possui limitações quanto à utilização dos sentidos, não conseguindo detectar expressões não-verbais, nem expressar emoções ou ter uma personalidade (TLILI et al., 2023). Luo et al. (2023, p. 12) sublinham que ainda é difícil para um *chatbot* imitar o ser humano, "ele pode simular um par, mas um muito sem graça, monótono e frio."

No contexto educacional, ressalta-se a importância das interações humanas profundas e da conexão emocional no processo educacional, as quais a IA não é capaz de replicar:

A educação é um processo interativo de longo prazo que necessita de interações duradouras e de alta qualidade, com conversas presenciais, conexões emocionais e trocas intelectuais entre professores e alunos sendo indispensáveis. (LUO et al., 2023, p. 12)

Ao adotar a linguagem natural nas interações com as máquinas e a depender do estilo da linguagem dos *chatbots*, observa-se um aumento da sensação de presença social. Hobert, Følstad e Law (2023) apontam que, em seu estudo, os participantes relataram que as formas de diálogo mais ativo elevaram, apesar de não ter sido de maneira significativa, a sensação de presença social. Porém, eles também indicaram que possivelmente o efeito do "Vale da Estranheza" poderia ser despertado nesses diálogos. Esse efeito consiste na sensação de estranheza, desconforto e repulsa que podem acontecer em vários tipos de interações entre

humanos e máquinas, causados pela dificuldade de distinção entre objetos humanóides e seres humanos reais. (CIECHANOWSKI et al., (2019)

Além da antropomorfização da IA, um assunto em debate é a substituição dos humanos por essa tecnologia. Desse modo, vale ressaltar que a utilização do *chatbot* não reduz a importância das funções dos professores nem da interação humana (JEON, LEE, 2023; CHAN, 2023). Ele pode auxiliá-los, mas não substituí-los (SMITH et al., 2023). No contexto da interação entre crianças e agentes conversacionais, destaca-se ainda mais a relevância da mediação por parte dos adultos (LUO et al., 2023).

Por isso, cabe aos educadores avaliar criticamente os recursos tecnológicos disponíveis e adaptá-los às necessidades específicas de seus contextos educacionais, levando em consideração a realidade de seus alunos (COOPER, 2023). Por essa razão, Pack e Maloney (2023) advertem que os educadores não devem encarar qualquer ferramenta baseada em IA como um mecanismo autônomo, mas sim como uma ferramenta que requer constante supervisão.

Assim, recomenda-se que o texto gerado pela IA seja analisado, revisado e refinado por um humano antes de sua utilização para fins educacionais. O professor deve, dessa maneira, se responsabilizar criticamente e ter o controle sobre suas escolhas, não delegando-as a sistemas de IA.

8.2.3 Dependência Excessiva em Sistemas de IA

Tan e Tan (2023) delineiam um cenário preocupante no contexto educacional, onde os seres humanos se limitam a pressionar algumas teclas e ninguém aprende nada, uma vez que a IA executa todas as tarefas. Nesse cenário, a IA não apenas cria a tarefa, mas também a conclui em nome do aluno, driblando os sistemas de detecção de plágio e até mesmo avalia a própria tarefa.

Chan (2023) sugere uma abordagem equilibrada na adoção da IA generativa, que reconheça tanto seus benefícios quanto suas limitações, evitando a dependência excessiva dessas ferramentas, as quais devem ser empregadas de forma complementar no ambiente educacional.

Para tarefas altamente complexas, como interações conversacionais mais elaboradas, orientação personalizada ou atividades de pesquisa científica, observa-se que a tecnologia ainda não consegue reproduzir de maneira eficaz os níveis de desempenho típicos do ser humano (PASSMORE, TEE, 2023).

Além de tudo, levando em consideração que os LLMs não têm experiências tangíveis no mundo real, Frick (2023) considera um equívoco profundo acreditar que são autoridades confiáveis sobre os assuntos referentes a esse mundo. Cooper (2023, p. 449) acrescenta que há um risco associado ao posicionamento do ChatGPT, por exemplo, "como a autoridade epistêmica definitiva, em que uma única verdade é presumida sem uma fundamentação adequada em evidências ou apresentada sem qualificações suficientes".

8.3 Literacia em IA

A literacia em IA refere-se à capacidade dos indivíduos compreenderem e utilizarem a IA de maneira eficaz, assim como ter discernimento para decidir quando recorrer ou não a essas tecnologias. Embora o diálogo em linguagem natural possa parecer mais intuitivo, a comunicação humano-máquina difere em algumas maneiras da comunicação entre humanos.

É evidente que há certa resistência à adoção de novas tecnologias. Porém, a falta de literacia em IA pode resultar em uso inadequado e limitado das ferramentas, afetando negativamente a qualidade da comunicação e do aprendizado. O progresso acelerado dessas ferramentas enfatiza ainda mais a importância de programas de treinamento para professores. Os alunos, por sua vez, devem ser orientados no uso crítico dos *bots* e incentivados a desenvolver habilidades holísticas, como pensamento crítico e criatividade. Na sequência, serão descritos os códigos relativos à literacia em IA.

8.3.1 Dificuldades na Adoção de Novas Tecnologias

A implementação de novas tecnologias é frequentemente acompanhada por manifestações de resistência, ansiedade e inquietações relacionadas à transição e incorporação dessas inovações. No estudo conduzido por Cardon et al. (2023), os participantes destacaram uma série de razões para a resistência à tecnologia, assim como as dificuldades enfrentadas durante o processo de adaptação. Eles consideraram, especialmente, os fatores: falta de competências prévias e indisponibilidade de tempo. Outros empecilhos referentes à adoção dessas

ferramentas foram identificados, como problemas sem soluções claras, a exemplo da distinção entre produções realizadas por estudantes e aquelas geradas por IA.

8.3.2 Alfabetização em IA

LUO et al. (2023, p. 14) observaram que o uso do ChatGPT requer um conjunto distinto de habilidades de alfabetização, abrangendo o que foi citado por um dos participantes de seu estudo: "[...] compreender a IA, saber como interagir com ela, discernir quando buscar sua ajuda e, especialmente, reconhecer quando não é necessário [...]" Os autores ressaltam a importância da alfabetização em IA tanto para crianças, quanto para pais e professores.

De acordo com Cardon et al. (2023), a alfabetização em Inteligência Artificial (IA) abrange quatro conjuntos distintos de habilidades. A primeira, denominada aplicação, refere-se à compreensão da utilização eficaz das ferramentas de IA generativa, assegurando sua relevância para as necessidades específicas do usuário. A segunda, autenticidade, implica na comunicação genuína e priorização da essência humana nas interações. Já a terceira, responsabilidade, está associada ao reconhecimento e à assunção das consequências decorrentes do conteúdo gerado por IA. Por fim, a agência envolve a retomada do controle pelos profissionais sobre suas próprias decisões.

8.3.3 Treinamento de Professores e Escolhas Pedagógicas

Com o rápido avanço da tecnologia, a dinamicidade e mutação são constantes no relacionamento entre seres humanos e IA. É fundamental que os professores adquiram domínio sobre essas ferramentas, bem como a confiança ao integrá-las às suas abordagens instrucionais (KOSTKA, TONCELLI, 2023). Dessa forma, "o ritmo acelerado das inovações em IA requer que os professores repensem e reimaginem suas filosofias de ensino" (TLILI et al., 2023, p. 19).

Jeon e Lee (2023) antecipam que à medida que as ferramentas de IA continuarem a evoluir, os papéis dos professores se tornarão cada vez mais diversificados e especializados. Eles alertam, ainda, que as novas ferramentas não necessariamente garantirão uma melhoria na qualidade do ensino. Em vez disso, essa qualidade dependerá do conhecimento pedagógico e do discernimento dos professores ao selecionar materiais adequados às necessidades individuais dos alunos em contextos instrucionais específicos.

A personalização do ensino mediada pela IA tem sido amplamente discutida, porém a dinâmica da sala de aula e integração da IA ainda é um desafio crucial a ser tratado. Jeon e Lee (2023) mencionam, ainda, que os professores têm uma percepção mais aguçada acerca do que sua turma precisa. Por isso, a adoção dos *chatbots*, assim como demais as decisões pedagógicas, devem levar em consideração vários fatores, dentre eles o "relacionamento entre o professor e os alunos, as necessidades e características individuais dos alunos e a dinâmica coletiva da sala de aula, com base na compreensão detalhada de seus alunos e contextos." (JEON, LEE, 2023, p. 13)

Dessa forma, a profissão docente seria redefinida no que se refere a três funções fundamentais: "orquestrar diferentes recursos com decisões pedagógicas de qualidade, tornar os alunos investigadores ativos e aumentar a consciência ética em relação à IA" (JEON, LEE, 2023, p. 16). Os autores frisam a necessidade de cursos de capacitação de professores direcionados especificamente para o ensino com auxílio de *chatbots* baseados em LLMs.

Alguns professores expressam preocupações sobre os desafios associados à alfabetização em IA dos alunos, tendo que transmitir essas informações sem possuir um preparo prévio adequado (CARDON et al., 2023). Ou seja, os professores estão enfrentando a necessidade de aprender a lidar com essas tecnologias enquanto também precisam ensiná-las aos alunos.

8.3.4 Treinamento de Alunos e Desenvolvimento de Habilidades Holísticas

É fundamental que "todos os alunos se tornem gestores de IA" (CARDON et al., 2023, p. 276), compreendendo suas aplicações nas atividades escolares e futuras carreiras. Esta compreensão deve abranger não apenas o potencial, mas também as limitações dessa ferramenta em diversos tipos de tarefas. Segundo Cardon et al. (2023), o ensino superior deveria preparar os alunos para o trabalho na Era da Inteligência Artificial, visando prepará-los adequadamente para as transições no mercado de trabalho.

Uma questão amplamente discutida ao adotar a IA na educação é o desenvolvimento de habilidades holísticas, como o pensamento crítico, a colaboração, a criatividade, a imaginação e a resolução de problemas. Luo et al. (2023) afirmam que é importante proporcionar às crianças oportunidades de desenvolverem a criatividade e de cultivarem seu próprio senso de imaginação antes de serem expostas à IA generativa.

Uma grande preocupação é o potencial impacto negativo do uso excessivo do ChatGPT sobre a capacidade de inovação e pensamento crítico dos estudantes, considerando o risco de a ferramenta ser utilizada como um atalho frequente para evitar esforços (TLILI et al., 2023). Sendo assim, um dos papéis dos professores é ensinar os estudantes a utilizar esses sistemas de maneira equilibrada e orientá-los na avaliação crítica da saída dos chatbots (FRICK, 2023; SU, LIN e LAI, 2023; ABDELGHANI et al., 2023; CARDON et al., 2023).

O ensino sobre a credibilidade, assim como a avaliação da validade e precisão das informações, tornou-se ainda mais essencial com o surgimento dos grandes modelos de linguagem. Esta habilidade, já relevante para a alfabetização digital, ganhou maior urgência, uma vez que a origem do conteúdo consumido se torna mais obscura (KOSTKA, TONCELLI, 2023).

Vasconcelos e Santos (2023) apresentam uma visão mais otimista, cujos resultados sugerem que tanto o ChatGPT quanto o Bing Chat promovem um ambiente de aprendizado mais envolvente, interativo e inclusivo. Além disso, eles afirmam que os *chatbots* mencionados impulsionam o desenvolvimento das habilidades holísticas, contribuindo para uma compreensão mais aprofundada do conteúdo educacional.

8.4 Engenharia de *Prompts*

A mensagem ou comando fornecido pelo usuário ao *chatbot*, conhecido como *prompt*, demanda uma formulação precisa e adequada para gerar a melhor resposta. Nesse tipo de interação dialógica, o indivíduo que elabora o *prompt* exerce um controle significativo sobre a resposta que será gerada a partir do vasto conjunto de informações contido nos *datasets* usados para treinar as *LLMs*. *Prompt Engineering* se refere à arte de criar comandos ou mensagens específicas para interagir com *chatbots* ou sistemas de IA, de modo a obter respostas precisas e relevantes.

Enquanto é comum comunicar-se de maneira ambígua com outros seres humanos, essa abordagem muda quando nos comunicamos com máquinas. Apesar de a comunicação em linguagem natural ser menos formal do que linguagens de programação, surge a necessidade de adotar uma abordagem mais precisa, estruturada e lógica na interação com máquinas, a fim de evitar ambiguidades. A formulação inadequada de *prompts* pode resultar em respostas imprecisas ou irrelevantes, afetando a eficácia da comunicação e reduzindo a utilidade das

interações com sistemas baseados em IA. Com isso, os desafios relacionados a esse tema, mencionados pelos artigos analisados, foram agrupados em duas subcategorias.

8.4.1 Habilidades de Questionamento e Estratégias de Prompt

"A ampla adoção do ChatGPT implica que essas ferramentas são fáceis de usar, pelo menos de maneira rudimentar" (CARDON et al., 2023, p. 277). Embora o diálogo baseado em *prompts* seja um método relativamente fácil de implementar, é necessário "[...] algum conhecimento sobre engenharia de *prompts* para gerar *outputs* relevantes" (ABDELGHANI et al., 2023, p. 30). Dessa forma, para refinar a interação com as LLMs e extrair seu potencial máximo, as habilidades de questionamento e de construção de *prompts* são fundamentais.

Um dos participantes do estudo de Luo et al. (2023, p.13) expôs que "[...] tudo se resume ao *prompt*", visto que o ChatGPT produz saídas baseadas na maneira como a pergunta é formulada. Assim, *prompts* cuidadosamente elaborados otimizam o desempenho do modelo de linguagem (VASCONCELOS, SANTOS, 2023). Jeon e Lee (2023) complementam essa ideia, salientando a importância de fazer perguntas precisas, fornecer instruções claras e concisas para obter as respostas desejadas.

Ademais, "a utilização de determinadas 'estratégias de *prompt*', que fornecem contexto adicional à pergunta, pode melhorar as chances de obter respostas e explicações corretas." (JALIL et al., 2023, p. 7) Compreende-se, portanto, que ao ser mais específico em relação ao comando, as respostas do ChatGPT podem ser aprimoradas significativamente e são minimizados os riscos de interpretações errôneas (PACK, MALONEY, 2023).

Por isso, é aconselhada a inserção de informações e contexto bem delimitados na *context window* — extensão do texto ou das interações anteriores que o modelo considera ao gerar uma resposta. No experimento de Smith et al. (2023, p. 5), por exemplo, "foi necessário fazer uma pergunta muito específica, com parâmetros bem definidos, para que o *chatbot* produzisse material prático relevante para a psiquiatria social".

Jeon e Lee (2023) também citaram que os professores, participantes de seu estudo, concordaram sobre a relevância de ensinar os alunos a desenvolver habilidades de questionamento. Em outras palavras, os alunos precisam desenvolver novas competências de investigação ativa e estilo de pensamento para a construção de *prompts*, já que isso

influenciará diretamente nos benefícios obtidos por eles em suas interações com os *chatbots* (TLILI et al., 2023). Essa abordagem promove a reflexão crítica dos alunos e contribui para o aperfeiçoamento da compreensão de diferentes conceitos, à medida que interagem ativamente com a ferramenta (VASCONCELOS, SANTOS, 2023).

A habilidade em elaborar comandos está intrinsecamente vinculada ao conceito de expertise técnica delineado por Cardon et al. (2023) no âmbito da literacia em IA. Portanto, a capacidade de formular adequadamente instruções, ou *prompts*, constitui uma parte das competências essenciais para compreender e aplicar eficazmente as ferramentas de IA, alinhando-as de maneira coerente com os objetivos das tarefas a serem executadas. Logo, a programação adequada dessas ferramentas requer, além do pensamento crítico, um nível específico de habilidade técnica.

Se você entrar no ChatGPT sem estar preparado - público, propósito, objetivos, etc. -, sua mensagem falhará. ... Como instrutores, nosso trabalho é ensinar os prós e os contras de usá-lo, de confiar demais nele, de não ser treinado em como projetar *prompts* efetivos para ele. (CARDON et al., 2023, p. 290).

Tlili et al. (2023) menciona, ainda, que os usuários podem se beneficiar de diferentes níveis de assistência à aprendizagem a depender da forma como estabelecem o diálogo e o estilo de sua interação com os *chatbots*.

8.4.2 Limitações das Interfaces

No que tange aos *chatbots* existentes, ainda existem limitações nas interfaces e, por consequência, nos diálogos estabelecidos. Rudolph, Tan e Tan (2023a) discutem as Limitações da *Context Window*, como a quantidade limitada de caracteres do *input* e as limitações dos formatos de conteúdo de mídia utilizados pelos agentes conversacionais, tendo como exemplo a incapacidade de desenhar diagramas e gráficos. Além disso, os *bots* também possuem limitações quanto à quantidade de caracteres do *output* (CHAUDHRY et al., 2023).

Um participante do estudo de Tlili et al. (2023) apontou a necessidade de mais funcionalidades no ChatGPT, como a capacidade de buscar informações específicas em conversas anteriores, o que considera conveniente para fins de pesquisa.

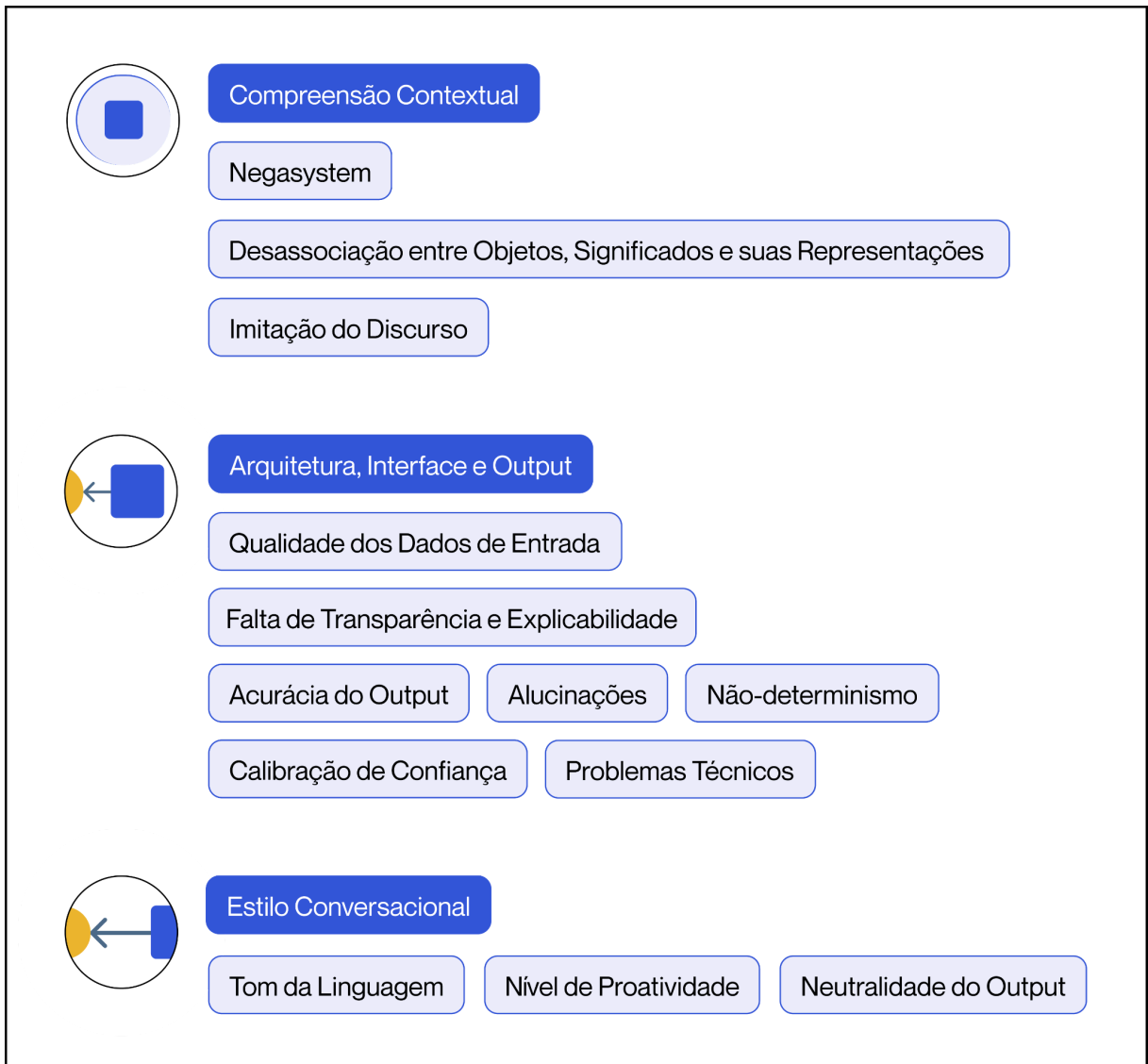


Figura 10 - Desafios Comunicacionais Relacionados ao LLM

8.5 Compreensão Contextual

O tema Compreensão Contextual refere-se à incapacidade das LLMs em capturar e interpretar o contexto humano em sua totalidade, incluindo suas nuances culturais, significados emocionais e outros aspectos complexos inerentes à interação humana. Essas limitações em compreender o mundo real, que é percebido por meio dos sentidos, resultam em algumas dificuldades para estabelecer uma comunicação efetiva com humanos, em termos de qualidade e profundidade, uma vez que os sistemas estão restritos às representações simbólicas da realidade em suas bases de dados.

8.5.1 *Negasystem*

Conceito introduzido por Frick (2023), *Negasystem* representa aquilo que os atuais sistemas de IA não conseguem acessar: a complexidade da cultura humana, composta por atividades únicas, sensações, percepções e emoções.

[...] emoções, e não estímulos cognitivos, servem como o arquiteto primário da mente". Cada sensação [...] também dá origem a um afeto ou emoção.... É essa codificação dupla da experiência que é fundamental para compreender como as emoções organizam as capacidades intelectuais [...] (GREENSPAN e BENDERLY, 1997 apud FRICK, 2023, p. 9).

Tais sistemas carecem dessas experiências diretas do mundo real, limitando-se aos registros criados por humanos, principalmente compostos por símbolos linguísticos, imagens audiovisuais e suas relações. Estes registros, incorporados aos sistemas de IA durante o treinamento, constituem meros "sinais que representam o mundo real e a cultura", sendo representações simbólicas e não os próprios objetos da experiência (FRICK, 2023, p. 1). Em suma, pode-se afirmar que os *chatbots* estão isolados do mundo real e que eles não são sencientes (do inglês, "*sentient*").

Essa conclusão nos abre espaço para discutir os limites das capacidades dos computadores e da tecnologia, particularmente em relação à "compreensão do contexto, empatia e análise para antecipar as emoções dos destinatários de nossas mensagens" (CARDON et al., 2023, p. 270).

8.5.2 *Desassociação entre Objetos, Significados e suas Representações*

Os atuais grandes modelos de linguagem não possuem conhecimento no sentido estrito da palavra.

É útil observar que o GPT-3, assim como seus antecessores e concorrentes, não possuem conhecimento no sentido estrito da palavra. Modelos de linguagem como esses não são projetados para armazenar ou recuperar fatos. Eles são "apenas bons em prever a próxima(s) palavra(s) na sequência" (COOPER, 2021 apud RUDOLPH, TAN, TAN, 2023a, p. 344).

A construção de modelos cognitivos confiáveis para textos complexos está intrinsecamente ligada à compreensão do funcionamento do mundo real (FRICK, 2023). Ao fazer referência ao princípio de Alfred Korzybski — "O mapa não é o território, e o nome não é a coisa nomeada" — o autor claramente destaca a limitação dos *chatbots* em criar modelos que se

baseiam não apenas em representações de objetos, mas sim na compreensão dos objetos reais, seus significados e das complexas relações entre eles.

[...] a aprendizagem profunda pode identificar objetos, mas não consegue compreender as relações entre eles e, muitas vezes, pode se enganar. [...] A aprendizagem profunda refere-se ao número de camadas em uma rede neural, não à profundidade de entendimento. (FRICK, 2023, p. 7)

8.5.3 Imitação do Discurso

A imitação do discurso acontece em paralelo aos códigos anteriores. Apesar de os *chatbots* não entenderem o conteúdo de suas afirmações (no sentido de compreensão conceitual entendido por seres humanos), eles refletem e reproduzem a linguagem técnica e sofisticada, bem como encadeamentos lógicos de argumentação, sem compreender o conteúdo em nível técnico (GREGORCIC, B., PENDRILL, 2023).

Gregorcic e Pendrill (2023) fornecem dois motivos pelos quais chegaram à conclusão de que há uma imitação do discurso: (1) a maneira como o *chatbot* é arquitetado, tendo sido treinado com um vasto conjunto de dados textuais, o que o capacita a fazer inferências estatísticas; e o fato de (2) as respostas fornecidas pela versão utilizada do ChatGPT não terem demonstrado bons níveis de conhecimento conceitual.

8.6 Arquitetura, Interface e Output

A arquitetura e *output* dos *chatbots* englobam desafios como qualidade dos dados disponíveis para os modelos de linguagem, a precisão das saídas produzidas pelo sistema e a ausência de transparência e explicabilidade no processo de tomada de decisão das redes neurais. Além disso, outros desafios relevantes incluem a confiabilidade e calibração dos LLMs, as chamadas "alucinações" e o não determinismo. Esses desafios podem acarretar respostas inadequadas, informações imprecisas e inconsistências nas respostas fornecidas.

As implicações decorrentes dessas questões são multifacetadas, englobando problemas como o aprendizado e disseminação de informações equivocadas, a falta de confiança nas interações entre humanos e IA, bem como dificuldades na compreensão e avaliação das respostas dos *chatbots*.

8.6.1 Qualidade dos Dados de Entrada

Em primeiro lugar, a qualidade de dados de entrada é crucial para garantir a qualidade do *output* do sistema. Como os grandes modelos de linguagem são treinados com dados em escala de *internet*, Frick (2023) comenta sobre a importância de filtrá-los, considerando que se as fontes utilizadas contiverem elementos que representem mentiras, desinformação, distorção da verdade, mitos, recomendações antiéticas, mensagens de ódio, entre outros, é provável que o resultado do sistema contenha esse tipo de conteúdo indesejável.

8.6.2 Falta de Transparência e Explicabilidade

Não há um entendimento completo do funcionamento interno de um sistema de IA, inclusive de suas redes neurais. Segundo o curso "*AI for Everyone*" (DeepLearning.AI, 2019), no contexto do treinamento de redes neurais, é estabelecida a relação entre um dado de entrada, representado por A, e sua respectiva saída, representada por B. Nesse processo, a IA é encarregada de realizar a predição do percurso necessário para transitar do input A ao output B.

Vale ressaltar que não são conhecidos os elementos intermediários ou as etapas específicas que a IA utiliza para realizar essa predição. Mesmo com ferramentas sofisticadas, especialistas têm dificuldade em entender por que redes neurais específicas tomam as decisões que tomam e como chegam a determinadas conclusões (FRICK, 2023; SMITH et al., 2023).

Este problema é agravado pelo fato de que os LLMs ainda carecem da capacidade de explicar o processo pelo qual chegam a essas conclusões. Em outras palavras, falta "explicabilidade" nos métodos utilizados pelas LLMs, levando à caracterização desses modelos como "caixas-pretas" (FRICK, 2023). Argumenta-se que uma IA responsável deveria se preocupar com a interpretabilidade e explicabilidade de suas conclusões (TLILI et al., 2023).

8.6.3 Acurácia do Output

É necessário esclarecer que o desenvolvimento dos LLMs estão ocorrendo de maneira notavelmente acelerada. Da transição da versão GPT-3 e GPT-3.5 para o GPT-4, observaram-se avanços substanciais em termos de qualidade e precisão dos *outputs* (PASSMORE, TEE, 2023). Diz-se que ChatGPT é eficaz em aumentar as chances de sucesso

educacional, proporcionando aos usuários (professores e alunos) conhecimento básico sobre vários tópicos (TLILI et al., 2023). No entanto, a depender do nível de dificuldade e natureza da tarefa, os *chatbots* demonstram desempenhos variados.

Em uma comparação sistemática entre o ChatGPT (GPT-3.5), ChatGPT (GPT-4), Bing Chat e Bard, Rudolph, Tan e Tan (2023b) realizaram uma avaliação multidisciplinar e encontraram diversas limitações dos sistemas. Os autores afirmaram que "atualmente não há alunos nota A e nem alunos nota B nesse grupo de bots, apesar de todas as alegações publicitárias e sensacionalistas" (RUDOLPH, TAN, TAN, p.377).

Por sua vez, Vasconcelos e Santos (2023) afirmaram que seus resultados em relação à qualidade das respostas do ChatGPT contrastaram com aqueles de Gregorcic e Pendrill (2023). Enquanto estes últimos concluíram que o ChatGPT não atendeu aos critérios necessários para ser usado como tutor de física, os primeiros sugeriram que o ChatGPT e o Bing Chat funcionam como excelentes "*objects-to-think-with*", ou seja, "ferramenta ou recurso que os aprendizes podem utilizar para explorar e desenvolver sua compreensão de um conceito ou domínio específico" (Papert, 1980 apud VASCONCELOS, SANTOS, 2023, p. 2).

Chaudhry et al. (2023) também encontraram bons níveis de acurácia dos *outputs*. Os autores concluíram que, dentro do contexto das avaliações de ensino superior testadas em seu experimento, o ChatGPT foi inteligente o suficiente para "escrever tarefas, analisar estudos de caso, desenvolver relatórios de projeto e fornecer soluções para problemas" (CHAUDHRY et al., 2023, p. 26). Com isso, o estudo constatou o desafio de aferir o desempenho dos alunos, que podem entregar tarefas inteiramente realizadas pelo *bot*.

A acurácia dos *outputs* em educação é particularmente relevante para garantir que os estudantes não aprendam com conteúdo incorreto ou tendencioso (HOBERT, FØLSTAD, LAW, 2023). Por produzir mensagens linguisticamente sofisticadas e de natureza convincente, há uma preocupação em relação a possíveis disseminações de informações incorretas, enganosas, sem sentido ou contraditórias (GREGORCIC, PENDRILL, 2023; SMITH et al., 2023; SU, LIN, LAI et al., 2023; PASSMORE, TEE, 2023; TLILI et al., 2023).

Passmore e Tee (2023) teorizam que as atualizações dos modelos de linguagem podem melhorar a precisão da saída, mas também aumentar a qualidade da falsificação, complicando a verificação das declarações e referências dos *chatbots*.

Ao abordar as causas das respostas incompletas ou imprecisas, as justificativas do ChatGPT são aparentemente arbitrárias, como falhas técnicas ou questões de formatação, atribuindo assim uma justificativa que não necessariamente reflete transparência ou veracidade (TLILI et al., 2023).

8.6.4 Alucinações

As alucinações são incidentes nos quais os modelos de linguagem produzem informações incorretas. É reconhecido o fenômeno das referências acadêmicas fictícias criadas durante interações com esses modelos. Rudolph, Tan e Tan (2023b) evidenciaram alucinações do Bard, que forneceu informações contrárias à narrativa de uma obra literária conhecida. Pack e Maloney (2023) ressaltaram que se os LLMs transmitirem conhecimento aos alunos, a presença de informações imprecisas torna-se um problema significativo, comprometendo a confiabilidade em relação ao conteúdo das mensagens recebidas.

8.6.5 Não-Determinismo

Outra questão levantada foi o aspecto não-determinístico das IAs generativas. Ao responder à pergunta "Como o não-determinismo do ChatGPT afeta sua capacidade de fornecer respostas e explicações corretas?", JALIL et al. (2023) encontraram que o ChatGPT pode apresentar respostas variadas em cada interação devido à natureza estocástica de seu método de amostragem. Assim, para a versão do ChatGPT avaliada, descobriu-se que algumas das respostas fornecidas pelo sistema diferem em precisão.

Os LLMs têm como característica a falta de replicabilidade, sendo que as respostas não podem ser reproduzidas de maneira exata (SMITH et al, 2023). No âmbito da avaliação de redações de estudantes, Su, Lin e Lai (2023) constataram que a cada vez que a redação do aluno foi submetida ao ChatGPT, este forneceu um feedback distinto. A falta de consistência restringe sua utilidade em avaliações e aconselha-se que não seja utilizado de maneira não supervisionada, embora possa ser usado como uma opinião alternativa (WIESER et al., 2023).

Uma implicação adicional decorrente dessa característica dos *bots* é a questão da equidade e imparcialidade no tratamento dispensado a todos os usuários. Tlili et al. (2023) identificaram em seus testes que, ao formularem o mesmo *prompt*, os usuários obtiveram respostas discrepantes, apresentando variações nos níveis de qualidade, organização e atualização das

informações fornecidas pelos *bots*. Isso levanta a questão fundamental de como assegurar um acesso equitativo e um tratamento imparcial para todos os usuários, sejam professores, alunos ou outros, com acesso a conteúdo de alta qualidade e atualizado.

8.6.6 Calibração de Confiança

Calibração de confiança se refere ao nível de confiança expresso pelo chatbot e sua correspondência com a precisão do resultado gerado. JALIL et al. (2023, p. 8) concluíram que a versão avaliada do ChatGPT era um "mau avaliador de sua própria correção" e que "para questões de teste de software, o ChatGPT não estava bem calibrado." Por exemplo, quando ChatGPT estava "altamente confiante" em sua resposta, descobriu-se que o output estava incorreto cerca de metade das vezes.

Na mesma linha, Gregorcic e Pendrill (2023) notaram uma discrepância entre o nível de confiança, a precisão das respostas e as habilidades linguísticas avançadas do ChatGPT. Eles observaram que, embora o chatbot utilize terminologia específica da física e crie argumentos logicamente organizados, aparentemente confiante em sua compreensão, persistem imprecisões e contradições em suas interações.

8.6.7 Problemas Técnicos

Esse código abrange as falhas, interrupções no fluxo da interação e limitações no funcionamento do *chatbot*. Tais problemas podem afetar a qualidade das interações, minando a confiança na utilidade e consistência dos sistemas. Rudolph, Tan e Tan (2023a, p. 348) mencionaram o aspecto de interrupção na comunicação: "Após fazer a pergunta no ChatGPT, a IA mostrou um alerta de erro de rede." Outra questão refere-se à limitação temporal dos modelos, como as versões do ChatGPT que possuem dados limitados sobre eventos mundiais até 2021. Considerando que algumas tarefas exigem informações atualizadas, isso se torna um desafio significativo (SMITH et al., 2023).

8.7 Estilo Conversacional

O estilo conversacional apresenta questões como a necessidade de adaptar a linguagem, tom e abordagem da interação do *chatbot* conforme os objetivos específicos e as necessidades individuais do usuário. Esse desafio impacta diretamente na eficácia do diálogo entre humano e IA, podendo influenciar negativamente o engajamento e provocar respostas afetivas

indesejadas nos usuários. Além disso, um estilo de interação inadequado pode afetar a percepção de autenticidade do *chatbot*, assim como limitar sua capacidade de oferecer suporte proativo, especialmente em contextos educacionais ou interativos.

8.7.1 Tom da Linguagem

O tom da linguagem também foi uma questão relevante mencionada, visto que o estilo conversacional deve estar alinhado com as expectativas e objetivos de uso. Expressões informais podem gerar uma percepção de inadequação ou incompatibilidade com contextos educacionais mais utilitários. Dessa maneira, as expectativas corretas podem ser definidas desde o início da interação (HOBERT, FØLSTAD, LAW, 2023).

8.7.2 Nível de Proatividade

Su, Lin e Lai (2023) mencionam que a capacidade generativa dos *chatbots* é reativa e depende de maneira significativa da qualidade da entrada fornecida pelo usuário. Embora o ChatGPT possa contribuir para o aspecto dialógico da escrita argumentativa ao apresentar aos alunos diferentes perspectivas, seu papel como colega virtual é limitado devido à natureza de sua capacidade generativa, condicionado a produzir conteúdo apenas quando solicitado. Nesse sentido, ele não é capaz de iniciar diálogos, oferecer orientação proativa ou auxiliar na estruturação do pensamento do aluno (SU, LIN, LAI, 2023).

Vasconcelos e Santos (2023) fornecem um exemplo de como formular o *prompt* para que o *chatbot* demonstre maiores níveis de proatividade, engajando o usuário em um diálogo com abordagem socrática:

Estou prestes a embarcar em um diálogo, no qual assumirei o papel de estudante, interagindo com você, ChatGPT, como um '*object-to-think-with*', com o objetivo principal de aprofundar e deliberar sobre conceitos físicos [ou qualquer tema escolhido]. Como um '*object-to-think-with*', solicito gentilmente o seu apoio para fornecer respostas envolventes e personalizadas às perguntas do estudante. É crucial que nossa interação adote uma abordagem socrática, incentivando o estudante a refletir sobre seu conhecimento existente, estimulando o pensamento crítico e auxiliando no aprimoramento da compreensão, ao invés de oferecer respostas prontas. Você está preparado para embarcar nesta jornada colaborativa e provocativa? (VASCONCELOS e SANTOS, 2023, p. 7)

8.7.3 Neutralidade do Output

Rudolph, Tan e Tan (2023a) descreveram o conteúdo gerado pelo ChatGPT como superficial, genérico e descritivo. Já Frick (2023, p. 5) definiu as respostas do Bard como "estéreis, porém razoáveis e práticas". Su, Lin e Lai (2023) complementam ao afirmar que os diferentes argumentos fornecidos pelo ChatGPT são superficiais. Eles são úteis como sugestões de ideias, mas não se apresentam como um conteúdo pronto para ser utilizado, por exemplo, na escrita.

Essas constatações são ecoadas pelos trabalhos de Passmore e Tee (2023) e Tlili et al. (2023), os quais indicam que as respostas do ChatGPT demonstraram ser vagas e genéricas, carecendo de exemplos ou detalhes específicos e apresentando um tom mecânico. Os argumentos fornecidos pelo *chatbot* foram descritivos e carentes de citações ou exploração detalhada do tema proposto.

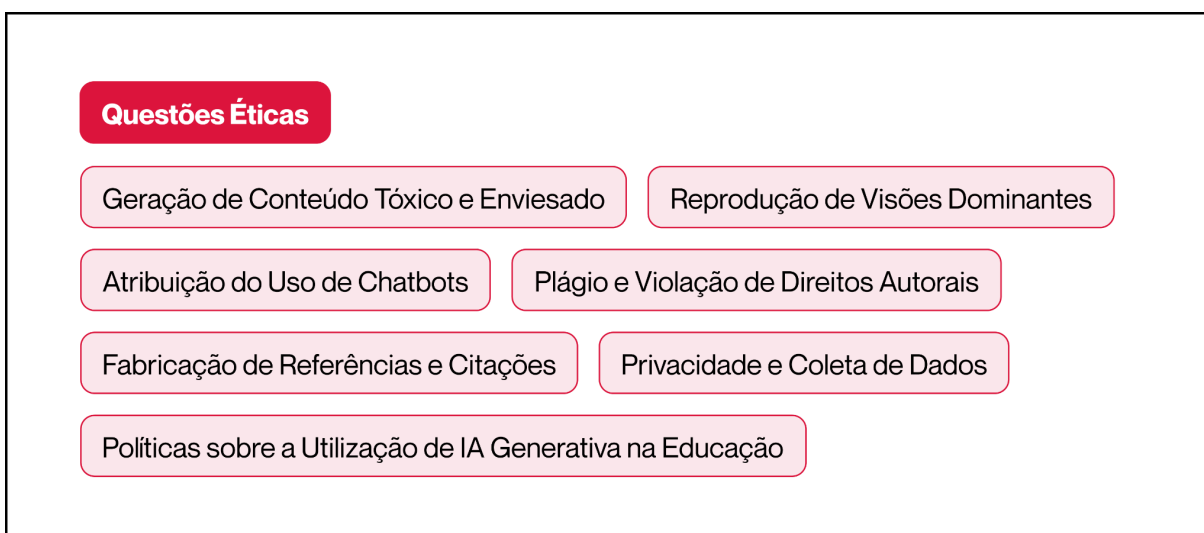


Figura 11 - Desafios Comunicacionais Relacionados a Questões Éticas

8.8 Questões Éticas

As questões éticas abrangem diversas preocupações, como a geração de conteúdo tendencioso ou reprodução de pontos de vista dominantes. Levanta-se também a questão da falta de atribuição do uso dos *bots*, dos potenciais casos de plágio e da ausência ou fabricação de citações e referências. Outra problemática é a coleta não regulamentada de dados dos usuários, sendo que muitas empresas que utilizam IA visam estabelecer uma estratégia de ciclo contínuo de coleta de dados, melhoria do produto e captação de mais usuários

(DeepLearning.AI, 2019). E, sobretudo, fala-se da necessidade de estabelecer políticas e diretrizes claras para a integração da IA na educação.

8.8.1 Geração de Conteúdo Tóxico e Enviesado

Em '*Generative AI with Large Language Models*' (DeepLearning.AI, Amazon Web Services, 2023), define-se toxicidade da seguinte maneira: "o LLM retorna respostas que podem ser potencialmente prejudiciais ou discriminatórias em relação a grupos específicos de pessoas ou a determinados atributos". Diversos estudos evidenciam que os LLMs podem gerar informações incorretas e são suscetíveis à propagar vieses (SMITH et al., 2023; VASCONCELOS E SANTOS, 2023; PACK, MALONEY, 2023; TLILI et al., 2023). Ademais, também há casos de geração de textos relacionados à violência, sexismo e racismo (SU, LIN, LAI, 2023).

O emprego de IAs, marcadas por esses "ruídos", no âmbito educacional pode acentuar várias formas de injustiça, como os vieses sistêmicos (AKGUN, GREENHOW, 2022 apud PACK, MALONEY, 2023). Adicionalmente, os estudantes podem, simplesmente, replicar conceitos incorretos, incompletos ou enviesados ao copiar e colar informações, o que torna desafiador o papel dos professores na identificação desses casos e mitigação dos problemas mencionados (VASCONCELOS, SANTOS, 2023).

8.8.2 Reprodução de Visões Dominantes

As diretrizes éticas estabelecidas pelos treinadores dos modelos, principalmente originários dos Estados Unidos, destacam uma tendência na ética do ChatGPT de refletir predominantemente os valores culturais ocidentais, negligenciando perspectivas de outras regiões (LUO et al., 2023; COOPER, 2023). Nesse contexto, Luo et al. (2023) alertam para as distinções culturais entre países, apontando possíveis limitações e preocupações resultantes dessa perspectiva culturalmente enviesada.

Além de sua visão limitada da verdade, a saída do LLM é fundamentada em considerar como verdadeiro o que é mais repetido (DARICS, POPPEL, 2023 apud COOPER, 2023). Dessa maneira, o questionamento sobre quais vozes são silenciadas pelo algoritmo e a identificação do viés do autor são questões cruciais para educadores e estudantes refletirem com cautela (COOPER, 2023).

8.8.3 Atribuição do Uso de Chatbots

No estudo conduzido por Jeon e Lee (2023), a atribuição na utilização de *chatbots* foi discutida, ressaltando a importância de alcançar um consenso sobre até que ponto os alunos podem recorrer ao ChatGPT. Eles enfatizaram a necessidade de encorajar o uso da ferramenta para aprimorar a compreensão, ao mesmo tempo promovendo um ambiente transparente em relação à atribuição de créditos aos *chatbots*.

Tlili et al. (2023) mencionaram as preocupações relacionadas à trapaça e Cardon et al. (2023), com a possibilidade de os alunos não assumirem a responsabilidade pelas informações geradas por *software* generativo, especialmente aquelas que são imprecisas e tendenciosas.

Um passo importante para a transparência do uso de *software* generativo é a conscientização dos alunos acerca da importância da integridade acadêmica e sobre as repercussões associadas a comportamentos inadequados (CHAUDHRY et al., 2023).

8.8.4 Plágio e Violação de Direitos Autorais

Em um primeiro momento, cabe pontuar que "As ideias e informações das pessoas estão sendo usadas pelo ChatGPT sem dar a elas crédito" (LUO et al., 2023, p. 12). Os LLMs coletam conteúdo e o reproduzem, não incluindo citações nem referências nas respostas fornecidas (RUDOLPH, TAN, TAN, 2023a; PASSMORE, TEE, 2023). Pack e Maloney (2023, p. 15) acrescentam que "o texto gerado por IA pode ser considerado como não original e, portanto, equivalente a plágio" e Passmore e Tee (2023, p. 9) questionam "a quem pertence este conteúdo?".

Isso traz à tona a problemática das violações de direitos autorais quando o texto produzido assemelha-se ou reproduz conteúdos preexistentes de outras fontes, sendo que essas fontes não são divulgadas ao usuário (KARIM, 2023 apud COOPER, 2023; SMITH et al., 2023). Por isso, Su, Lin e Lai (2023) propõem uma redefinição do conceito de plágio, argumentando que a reprodução de textos gerados por LLMs sem a devida atribuição deveria ser considerada dentro desse escopo. Visto que essas ferramentas agregam cada vez mais dados em seus espaços informacionais, torna-se crucial considerar como os autores e editores do conteúdo consumido podem ser remunerados por seu trabalho (PASSMORE, TEE, 2023).

A falta de referências gera ramificações desse problema, como a incerteza sobre a precisão das informações, a falta de evidências para apoiar os *outputs* (COOPER, 2023) e a incapacidade de verificar a precisão das informações fornecidas (LUO et al., 2023). Além disso, os *chatbots* ainda são indetectáveis para a maioria dos *softwares* antiplágio (CHAUDHRY et al., 2023), já que o *output* é único, refletindo a maneira como a pergunta é formulada.

8.8.5 Fabricação de Referências e Citações

Os usuários frequentemente se deparam com erros ou falsificações em referências e citações. Por exemplo, no estudo de Frick (2023), o Bard não fornecia as fontes das informações, a menos que fossem explicitamente solicitadas. Além disso, o *chatbot* apresentou erros em muitas das referências, como nomes de autores incorretos, datas imprecisas e títulos incompletos ou incorretos e assim sucessivamente. A incapacidade de fornecer referências, assim como a fabricação de referências e citações também foram observadas pelo estudo de Rudolph, Tan e Tan (2023a), Rudolph, Tan e Tan (2023b) e por Passmore e Tee (2023). Tlili et al. (2023) destacam o argumento dos participantes de seu estudo, os quais relataram que os dados de saída do ChatGPT parecem mais uma opinião quando não são acompanhados por referências adequadas.

8.8.6 Privacidade e Coleta de Dados

Por fim, outra questão ética importante é a privacidade e coleta de dados. Luo et al. (2023, p. 12) mencionaram que "Não está claro como o ChatGPT armazena e processa as interações do usuário, o que levanta preocupações sobre privacidade, segurança de dados e ética."

Os participantes do estudo conduzido por Tlili et al. (2023) demonstraram inquietações com a exposição de seus dados pessoais e demográficos durante as interações com o ChatGPT. Um ponto preocupante foi que, durante essas interações, o *bot* negou a coleta de dados das conversas dos usuários. No entanto, não há clareza quanto ao armazenamento e uso dessas conversas. Embora possam contribuir para a melhoria do sistema, é crucial esclarecer como esses dados são tratados. Especialmente os usuários mais jovens devem ser alertados e instruídos sobre essas questões.

8.8.7 Políticas sobre a Utilização de IA Generativa na Educação

No estudo de Cardon et al. (2023), educadores expressaram a necessidade de diretrizes centralizadas e critérios definidos para o emprego da Inteligência Artificial generativa. Muitos também compartilharam a opinião de que é fundamental estabelecer regras e restrições para o uso das ferramentas de escrita automatizada pelos alunos, porém demonstraram preocupação quanto à formulação de políticas demasiadamente rígidas, precipitadas e de curto prazo. Há também aqueles, cujos argumentos envolvem a integridade acadêmica e o desenvolvimento dos estudantes, que advogam pela proibição do uso da IA generativa,

Chan (2023) fornece algumas diretrizes para a elaboração dessas políticas, agrupando-as em três dimensões: a dimensão pedagógica, que trata dos aspectos de ensino e aprendizagem; a dimensão de governança, relacionada a questões da alta administração; e a dimensão operacional, que discorre sobre a implementação prática da IA em ambientes universitários. A autora propõe que recomendações para criação de políticas sejam o primeiro passo para a criação de planos de ação para mitigar as limitações e desafios da utilização de *software* generativo.

É evidente que diversas políticas redigidas em 2022 ou em anos anteriores estão defasadas e poderiam ser aprimoradas por meio de revisões que abordam explicitamente o conteúdo gerado por IA e estabeleçam parâmetros claros para seu uso aceitável dentro da estrutura organizacional (PASSMORE; TEE, 2023).

9. DISCUSSÃO

Os grandes modelos de linguagem estão ganhando uma presença cada vez maior nas dinâmicas dos ambientes educacionais. Esta revisão sistemática possibilitou o mapeamento de desafios de comunicação identificados nesse cenário, analisando um total de 22 artigos no domínio da educação.

Os seis artigos adicionais selecionados para esta revisão foram considerados relevantes no âmbito da comunicação entre humanos e inteligência artificial. Estes artigos serão analisados detalhadamente nesta seção, visando estabelecer um diálogo com os resultados obtidos. O título de cada subseção a seguir corresponde ao título do artigo analisado.

É relevante salientar que estes artigos correspondem a contextos não estritamente educacionais ou comunicacionais e que discutem questões mais técnicas da interação com a IA. Neste sentido, a intenção foi conectar essas pesquisas que, apesar de situadas no campo da computação, foram agrupadas na categoria de "comunicação", dada a escassez de artigos específicos na área da comunicação que abordem esse tema.

9.1 Prompting Large Language Models with The Socratic Method (CHANG, 2023)

O cerne do estudo de Chang (2023) está na **engenharia de prompts**. O autor cita que ainda é complexa a criação de comandos eficazes, que demanda um entendimento profundo acerca da interação entre o LLM e o *prompt*. É preciso contextualizar o modelo de maneira precisa e expressar de forma clara a intenção do usuário para obter respostas de desempenho ótimo.

Chang (2023) propõe uma elaboração rigorosa de *templates* de *prompts*, com base no método socrático, que utiliza o raciocínio indutivo, dedutivo e abduutivo, para orientar e aprimorar os *outputs* obtidos. Os métodos implementados são os seguintes: definição, elenchus, dialética, maiêutica e técnicas de raciocínio contrafactual.

Ao empregar essas estratégias, o usuário é capaz de extrair do modelo de linguagem um papel ativo na construção das **habilidades holísticas** do próprio usuário. Como, por exemplo, o **pensamento crítico**, por meio da elicitación de informações precisas e relevantes; e o **pensamento criativo**, com a estimulação da imaginação dos usuários. Portanto, as abordagens de definição, elenchus e dialética aprimoram a precisão, consistência e coerência dos *prompts* elaborados e, também auxiliam na avaliação dos resultados gerados pelo *chatbot*. Já a aplicação da maiêutica e do raciocínio contrafactual contribui para estimular a imaginação, fomentando a autodescoberta e a escrita criativa, ao explorar cenários alternativos para determinados eventos (CHANG, 2023).

9.2 How General-Purpose is a Language Model? Usefulness and Safety with Human Prompters in the Wild (CASARES et al., 2022)

Na comunicação com *bots* de IA, há duas principais etapas: a **formulação de comandos em forma de prompts**, de forma apropriada para direcionar ou induzir o sistema a completar os estímulos de acordo com a intenção do usuário (CASARES et al., 2022). Os autores relembram também o caráter **não-determinístico** dos sistemas, sendo que pequenas variações

do *prompt* podem tornar os resultados muito piores. A segunda etapa é a decodificação, que consiste na interpretação e extração da parte relevante da resposta gerada pelo sistema.

Casares et al. (2022) destacam a lacuna existente nos relatórios de desempenho das LLMs, enfatizando a ausência de consideração pelos contextos reais de interação com os *bots*. Destacam a importância de testá-los em uma "situação média", visando compreender a interação humana típica com esses sistemas, considerando tanto os possíveis benefícios quanto os riscos associados. Nesse cenário, a validade ecológica emerge como palavra-chave, pois avalia a adaptação das LLMs aos diversos contextos humanos.

Seguindo esse raciocínio, os autores propõem um novo *framework* teórico de avaliação. Essa avaliação compreende a mensuração da utilidade, ou seja, a capacidade dos humanos de **formular *prompts*** e decodificar as respostas do modelo de maneira mais rápida do que conseguiriam resolver a tarefa por si mesmos. Para alcançar isso, é necessário medir a diferença entre os custos da utilização do modelo e o custo que o humano teria sem o modelo. Os custos da utilização do modelo englobam o esforço humano necessário para criar *prompts* adequados, desenvolver estratégias de codificação e decodificação, bem como o esforço humano associado à aplicação dessas estratégias na redação do *prompt* e na extração dos resultados. Isso inclui também o custo humano de validar ou descartar a solução fornecida pelo modelo.

Constata-se que o uso desses modelos ainda exige um esforço significativo por parte do humano comum para gerar bons *prompts* (CASARES et al., 2022). A **literacia em IA** poderia, a longo prazo, diminuir os esforços e o custo envolvido na utilização dos *chatbots*.

A avaliação proposta também envolve as **falhas no modelo**, seja por falta de capacidade ou por falta de entendimento do comando emitido pelo usuário. O artigo concluiu em seu experimento que, para diversas tarefas, exceto a escrita generativa, seria mais rápido se o usuário resolvesse a tarefa sem o GPT-3.

9.3 Don't Just Tell Me, Ask Me: AI Systems That Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy Over Causal Ai Explanations (DANRY et al., 2022)

Os resultados apontam para uma busca por equilíbrio na utilização da Inteligência Artificial. Esse equilíbrio visa evitar uma **dependência excessiva na ferramenta**, reconhecendo a importância da agência e da essência humana nas tarefas e interações. A preocupação reside na possibilidade de uma inclinação desmedida para o uso da IA em detrimento do **desenvolvimento de habilidades holísticas**, como o pensamento crítico. Tal questão foi discutida por Danry et al. (2023), destacando também a derivada **percepção de autoridade** dessas ferramentas, levando à delegação de responsabilidades que deveriam ser exercidas pelos humanos.

Este comportamento suscita preocupações sobre a susceptibilidade dos usuários às potenciais fraquezas da IA, como **imprecisões nos resultados gerados**, a **propagação de conteúdo enviesado e tóxico**, assim como a **perpetuação de visões dominantes** inerentes ao modo de produção desses sistemas.

Uma estratégia sugerida por Danry et al. (2023) como tentativa de mitigar esse problema foi uma abordagem diferenciada no design dos *chatbots*. Em vez de perpetuar o **ciclo de dependência**, os próprios *chatbots* podem ser programados para auxiliar os seres humanos no **desenvolvimento do pensamento crítico**. Isso implicaria o aumento da carga cognitiva no que tange à aferição da validade lógica dos *outputs*, bem como a cultivação da habilidade de discernir quando confiar ou não nos grandes modelos de linguagem.

Essa estrutura de apoio ao pensamento permite aos usuários avaliar e tomar decisões independentes, em oposição a aceitar passivamente as informações da IA sem questionamento. A proposição vai ao encontro do que foi proposto por Vasconcelos e Santos (2023), que consideraram as ferramentas de IA como um suporte valioso para explorar e promover o desenvolvimento das habilidades cognitivas holísticas.

A sugestão prática de Danry et al. (2023) para concretizar essa abordagem é a implementação de perguntas estruturadas seguindo o método socrático, mas não por parte do usuário, e sim do modelo de linguagem. Essa estratégia projeta os *chatbots* para serem **mais proativos** e oferecerem uma alternativa à explicação causal usualmente retornada por esses sistemas, já

que esse modelo de explicação, apoiado por linguagem sofisticada e confiante, reduz as oportunidades de questionamento por parte dos usuários. Portanto, é essencial que a IA seja capaz de discernir quando perguntar e quando informar, sobretudo em contextos de aprendizagem.

9.4 Co-Writing with Opinionated Language Models Affects Users' Views (JAKESCH ET AL., 2023)

Jakesch et al. (2023) revisitam a questão **da toxicidade e da geração de conteúdo tendencioso** nos modelos de linguagem, incluindo a reprodução de estereótipos, preconceitos e a **reprodução de perspectivas dominantes**, evidenciando a valorização de certos valores culturais em detrimento de outros. Os autores também afirmam que os modelos apresentam riscos sociais e aprendem esse tipo de comportamento a partir dos **dados de treinamento**. Como resultado, esses modelos têm um **desempenho superior em idiomas e contextos nos quais foram treinados**, especialmente em inglês ou mandarim (JAKESCH et al., 2023). Os autores mencionam que o GPT-3, por exemplo, revelou estar mais alinhado aos valores predominantes nos Estados Unidos.

No experimento realizado, ao utilizar os *chatbots* como co-autores na produção de texto, os pesquisadores identificaram uma notável influência tanto na redação dos participantes quanto em suas atitudes em relação ao tema proposto durante o experimento. Houve uma tendência dos participantes em adotar a opinião favorecida pelo modelo, refletindo-se tanto na maneira como escreveram quanto em suas posturas em relação ao assunto. É revelado, também, no estudo a facilidade da construção de modelos altamente opinativos.

Embora o estudo tenha se concentrado em um contexto específico, ele suscita preocupações mais amplas sobre o uso de modelos de linguagem opinativos e seu potencial de influenciar tanto indivíduos quanto populações inteiras. O artigo introduz o conceito dessa influência como "persuasão latente", referindo-se a uma abordagem indireta e sutil, cuja intenção é moldar as atitudes e comportamentos das pessoas sem que elas percebam conscientemente esse processo. Esta influência é ainda mais amplificada pelo fato de os participantes atribuírem um alto nível de expertise ao assistente, possivelmente **considerando-o uma autoridade**, levando à aceitação dos conselhos do algoritmo, mesmo em situações simples quando está claramente equivocado.

Essas preocupações também se estendem a modelos similares integrados em outras aplicações, como teclados preditivos e assistentes de voz (JAKESCH et al., 2023). Com determinadas opiniões e perspectivas impregnadas na tecnologia, destaca-se, neste ponto, a importância de reavaliar a **dependência excessiva em sistemas de IA**, conceituada no artigo como "viés de automação" – a inclinação para confiar mais nos resultados originados de sistemas tecnológicos do que em nossa própria reflexão ou experiência.

Portanto, torna-se essencial supervisionar os modelos de linguagem que podem ser intencionalmente ou não calibrados de maneira inadequada e, paralelamente, promover o **desenvolvimento de habilidades holísticas** para fomentar a conscientização sobre esses possíveis tipos de influências.

9.5 “I Think You Might Like This”: Exploring Effects of Confidence Signal Patterns on Trust in and Reliance on Conversational Recommender Systems (RADENSKY et al., 2023)

O estudo conduzido por Radensky et al. (2023) centrou-se na investigação do impacto dos padrões de sinais de confiança emitidos pelos *chatbots*. O experimento envolveu a utilização de frases com diferentes níveis de confiança para representar a possibilidade, probabilidade ou certeza de os usuários apreciarem as músicas recomendadas pelo sistema. Os padrões de confiança testados foram divididos em três categorias: estático – em que a confiança era mantida constantemente alta ou constantemente baixa; aleatório – correspondendo ou não com o nível de confiança real do sistema; e preciso – refletindo com precisão o nível de confiança do sistema em suas afirmações.

Os indicadores de confiança refletem a **calibração de confiança** do sistema. Conforme delineado por Radensky et al. (2023), o propósito dos indicadores de confiança precisos não é incentivar um aumento excessivo na confiança nos sistemas, o que figuraria a **dependência excessiva nos sistemas**, mas sim auxiliar os usuários a calibrarem suas expectativas para cada recomendação. Esta questão pode ser especialmente relevante no domínio educacional, onde os sistemas frequentemente sugerem materiais de leitura, teorias ou abordagens para solucionar problemas que se beneficiariam de uma calibração apropriada e transparência.

O estudo constatou que a precisão dos sinais de confiança não gera dependência excessiva, mas sim uma possível subdependência. Dessa maneira, a transparência reduz as expectativas

exageradas e aumenta compreensão dos usuários em relação a falhas do sistema à medida que o modelo incorpora mais dados e aprende. Torna-se evidente que os participantes almejam a transparência dos sistemas e expressam a necessidade de explicações claras para recomendações fornecidas pela IA. Ademais, os autores identificaram que as abordagens para comunicar a confiança do sistema resultam em distintos níveis percebidos de competência, benevolência e **antropomorfismo** por parte dos usuários.

9.6 Prompting Meaning: A Hermeneutic Approach to Optimising Prompt Engineering with Chatgpt (HENRICKSON, MEROÑO-PEÑUELA, 2023)

Henrickson e Meroño-Peñuela (2023) exploram **estratégias de otimização de prompts** destinadas a alcançar *outputs* ótimos em termos de valor hermenêutico – ou seja, interpretações e significados alinhados com as circunstâncias contextuais. Os autores afirmam que a **engenharia de prompts** possui diversos níveis de complexidade e que o formato de 'bate-papo' permite que o usuário formule seus comandos de maneira iterativa para alcançar a resposta almejada.

Assim, a comunicação com modelos de linguagem tende a ser controlada e direcionada pelo emissor do *prompt*, sendo que, quanto maior domínio e especificidade do *prompt*, mais precisas são as respostas (HENRICKSON, MEROÑO-PEÑUELA, 2023). Pela forma como estão programados, os modelos de linguagem priorizam a precisão, obediência e concordância em detrimento do pensamento especulativo e criativo. No entanto, *prompts* formulados de maneira vaga, induzem a saídas caracterizadas pela **neutralidade**, gerando poucos *insights* e oferecendo respostas genéricas, triviais, repetitivas e carentes de originalidade.

Outra questão abordada é o fato de os LLMs espelharem perspectivas humanas, que não podem ser reduzidas a números ou *tokens* - tanto os humanos que os concebem quanto os que interagem com eles são refletidos nesses sistemas, moldando suas bases de dados. Portanto, o sistema não é neutro; é programado por humanos com perspectivas subjetivas e alimentado por dados também gerados por humanos, o que destaca a importância da **qualidade e abrangência dos dados inseridos** nesses sistemas.

Isto posto, o ChatGPT não funciona como uma nova forma de vida, ao contrário das **percepções antropomórficas**. Como apontado nos resultados, ele concretiza uma nova forma de colaboração social inovadora (LANIER, 2023 apud FRICK, 2023, p. 9), que ainda está na

posição de ilusão de inteligência, sustentada por trabalho humano. Deve-se considerar também quem está sendo incluído ou excluído dos conjuntos de dados desses sistemas.

Henrickson e Meroño-Peñuela (2023) apontam que a falha em reconhecer os *chatbots* como reflexos dos humanos pode derivar da falta de referências literárias explícitas que informem suas saídas, uma forma de revisão de literatura mal referenciada, porém sintaticamente original, formulando um "**plágio** de alta tecnologia" (HENRICKSON e MEROÑO-PEÑUELA, 2023).

Os autores, assim como Danry et al. (2023), interpretam a **falta de proatividade** das LLMs como uma lacuna. Essa falta de subjetividade e hermenêutica da interação, atribuída à ausência de questionamento proativo nos diálogos, prejudica a construção de uma interação bilateral significativa. Essa troca só ocorre quando os modelos de linguagem são devidamente programados através da **engenharia de prompts**, conferindo aos usuários a responsabilidade pela construção do significado na interação.

Finalmente, a dualidade presente nas LLMs revela a influência das perspectivas humanas inculcadas nos dados, mas também aponta para um extenso percurso desde a entrada das palavras, convertidas em tokens, até a saída em forma de conclusões no *output*. Isso acarreta uma possível perda de significado ao longo desse processo. Nesse contexto, evidenciam-se deficiências dos *chatbots* quanto à compreensão textual, em seu **negasystem** e a **desassociação entre objetos, seus significados e suas representações**.

10. CONCLUSÃO, LIMITAÇÕES E TRABALHOS FUTUROS

A comunicação entre humanos e máquinas está passando por uma transformação significativa. A utilização da linguagem natural com as máquinas difere daquela utilizada entre humanos, uma vez que demanda um nível mais elevado de precisão. Enquanto os humanos estão familiarizados com a ambiguidade e as múltiplas interpretações de eventos, as máquinas estão isoladas desses contextos. Para potencializar os benefícios das interações com a IA no contexto da educação, é necessário desenvolver proficiência na comunicação dialógica com sistemas computacionais, em que a resposta do modelo de linguagem depende amplamente dos comandos emitidos. Contudo, emerge a questão: devemos adaptar os *chatbots* para lidar de maneira mais efetiva com essas ambiguidades, garantindo maior

intuitividade na interação com as ferramentas, ou é imperativo que nos adaptemos e compreendamos essa nova forma de interação?

Talvez a resposta esteja em um meio termo, o que nos leva à importância de evitar tanto o determinismo tecnológico quanto o ludismo (RUDOLPH, 2018 apud RUDOLPH, TAN, TAN, 2023a) ao considerar o papel da IA na educação. Cabe, neste momento, ir além dos pensamentos dicotômicos que idealizam ou demonizam a tecnologia e que apenas retardam o desenvolvimento de estratégias, baseadas em raciocínio crítico, para saber lidar com elas. Dessa maneira, é interessante vislumbrar diferentes estratégias de colaboração humano-IA (TLILI et al., 2023) e, para isso, os resultados deste trabalho apontam que são cruciais a alfabetização em IA e o estabelecimento de políticas bem estruturadas para guiar a utilização da IA nas práticas educacionais.

A implementação da IA na educação pode ser interpretada como um grande alerta sobre os problemas preexistentes nesse campo. A capacidade da IA em buscar informações em seu banco de dados, juntamente com a tendência dos estudantes em utilizá-las como atalho para a realização de tarefas e avaliações, coloca em cheque metodologias de ensino que se limitam à memorização de informações. Isso instiga uma necessidade premente de reformular a abordagem pedagógica e a filosofia de ensino visando guiar os alunos para além da mera retenção de fatos, direcionando-os a um processamento cognitivo mais profundo e ao desenvolvimento de habilidades holísticas. Tal direcionamento alinha-se com os princípios da Taxonomia Revisada de Bloom, que advoga pela promoção das habilidades de análise, avaliação e criação (IOWA STATE UNIVERSITY, 2022).

Este estudo apresenta limitações que merecem consideração. Primeiramente, destaca-se o viés da pesquisadora, que pode influenciar a interpretação dos resultados. Além disso, a rápida evolução e atualizações frequentes nos modelos de linguagem de sistemas de IA podem tornar obsoletos certos aspectos das análises, limitando sua relevância a longo prazo. Por fim, destaca-se a carência de estudos provenientes especificamente do campo da comunicação no contexto analisado.

Para trabalhos futuros, sugere-se um mapeamento das potenciais soluções para os desafios identificados, a condução de estudos longitudinais para avaliar os impactos a longo prazo da presença da IA na educação e a identificação dos pontos-chave da alfabetização em IA,

examinando de forma aprofundada quais competências são essenciais para potencializar a colaboração humano-IA.

11. REFERÊNCIAS

ABDELGHANI, R. et al. GPT-3-Driven Pedagogical Agents to Train Children's Curious Question-Asking Skills. **International Journal of Artificial Intelligence in Education**, 30 jun. 2023.

ASHMAN, M. **Communication Models – Introduction to Professional Communications**.

Disponível em:

<<https://pressbooks.bccampus.ca/professionalcomms/chapter/3-2-the-communication-process-communication-in-the-real-world-an-introduction-to-communication-studies/>>. Acesso em: 1 dez. 2023.

BOLAND, A.; CHERRY, M ; DICKSON, Rumona. **Doing a Systematic Review: a student's Guide**. 2. ed. London: Sage, 2017.

BRAUN, V. ; CLARKE, V. **Using Thematic Analysis in Psychology**. *Qualitative Research in Psychology*, v. 3, n. 2, p. 77–101, 2006.

CARDON, P. W. et al. The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age. **Business and Professional Communication Quarterly**, v. 86, n. 3, p. 232949062311765-232949062311765, 30 maio 2023.

CARMICHAEL, L. **Two essays on user experience with chatbots**. 2022. 91 f. Dissertação (Maîtrise ès sciences en gestion, M. Sc.) – HEC Montréal, Montréal. Disponível em: https://biblos.hec.ca/biblio/memoires/carmichael_laurie_m2022.pdf. Acesso em: abril/2023.

CASARES, P. A. M. et al. How General-Purpose Is a Language Model? Usefulness and Safety with Human Prompters in the Wild. **Proceedings of the AAAI Conference on Artificial Intelligence**, v. 36, n. 5, p. 5295–5303, 28 jun. 2022.

CHAN, C. A Comprehensive AI Policy Education Framework for University Teaching and Learning. **International Journal of Educational Technology in Higher Education**, v. 20, n. 1, 7 jul. 2023.

CHANG, E. Y. **Prompting Large Language Models with the Socratic Method**. 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), p. 0351-0360. Las Vegas, NV, USA: IEEE, 2023.

CHAUDHRY, S. I. et al. Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT — a Case Study. *Cogent Education*, v. 10, n. 1, 14 maio 2023.

CHIU, T. K. F. et al. Systematic Literature Review on Opportunities, Challenges, and Future Research Recommendations of Artificial Intelligence in Education. **Computers and Education: Artificial Intelligence**, v. 4, p. 100118, 2023.

CIECHANOWSKI, L. et al. In the Shades of the Uncanny valley: an Experimental Study of Human–chatbot Interaction. **Future Generation Computer Systems**, v. 92, p. 539–548, mar. 2019.

COOPER, G. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. **Journal of Science Education and Technology**, v. 32, p. 444–452, 22 mar. 2023.

DAĞLI, M. **Designing for Trust**: Exploring Trust and Collaboration in Conversational Agents for E-commerce. 2018. 127 f. Dissertação (Master of Design in Interaction Design) – School of Design, Carnegie Mellon University, Pittsburgh. Disponível em: https://kilthub.cmu.edu/articles/thesis/Designing_for_Trust/7772492. Acesso em: abril/2023.

DANRY, V. et al. **Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI Explanations**. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). New York, NY, USA: Association for Computing Machinery, 19 abr. 2023.

DeepLearning.AI; Amazon Web Services. **Generative AI with Large Language Models**. Disponível em: <<https://www.coursera.org/learn/generative-ai-with-llms>>. Acesso em: 15 out. 2023.

DeepLearning.AI. **AI for Everyone**. Disponível em:

<<https://www.coursera.org/learn/ai-for-everyone>>. Acesso em: 2 dez. 2023.

DERMEVAL, D. et al. Authoring Tools for Designing Intelligent Tutoring Systems: a Systematic Review of the Literature. **International Journal of Artificial Intelligence in Education**, v. 28, n. 3, p. 336–384, 31 out. 2017.

FØLSTAD, A.; BRANDTZÆG, P. B. Chatbots and the new world of HCI. **Interactions**, v.24, n.4, p.38–42, jul./ago. 2017. DOI: 10.1145/3085558. Disponível em: <https://dl.acm.org/doi/10.1145/3085558>. Acesso em: abril/2023.

FONSECA, J. **Metodologia da Pesquisa Científica**. Universidade Estadual do Ceará, 2002.

FRICK, T. W. Are We Dupes? Limitations of AI Systems: What Should Educators Do with Them? **TechTrends**, v. 67, n. 5, 21 set. 2023.

GIL, A. C. **Métodos E Técnicas De Pesquisa Social**. São Paulo: Atlas, 2008.

GOUGH, D.; OLIVER, S. ; THOMAS, J. An introduction to systematic reviews. 2. ed. London ; Los Angeles: Sage Publications Ltd, 2017.

GRAESSER, A. C.; LI, H. ; FORSYTH, C. Learning by Communicating in Natural Language With Conversational Agents. **Current Directions in Psychological Science**, v. 23, n. 5, p. 374–380, 2014.

GREGORCIC, B.; PENDRILL, A.M. ChatGPT and the Frustrated Socrates. **Physics Education**, v. 58, n. 3, p. 035021–035021, 22 mar. 2023.

GRIOL, D.; CALLEJAS, Z. Conversational Agents as Learning Facilitators: Experiences With a Mobile Multimodal Dialogue System Architecture. In: **Formative Assessment, Learning Data Analytics and Gamification**. [s.l.] Elsevier, 2016.

GUNNING, David; STEFIK, Mark; CHOI, Jaesik; et al. XAI—Explainable artificial intelligence. *Science Robotics*, v. 4, n. 37, p. eaay7120, 2019. Disponível em: <https://robotics.sciencemag.org/content/4/37/eaay7120.full>. Acesso em: jun/2023.

HARARI, Y. Reboot for the AI Revolution. **Nature**, v. 550, p. 324–327, 2017. Disponível em: <https://www.nature.com/articles/550324a>. Acesso em: jun/2023.

HASHANA, A. et al. Deep Learning in ChatGPT - A Survey. In: **7th International Conference on Trends in Electronics and Informatics (ICOEI)**. 2023. Tirunelveli, India. IEEE. Disponível em: <https://ieeexplore.ieee.org/document/10125852>. Acesso em: jun/2023.

HE, Z. From ELIZA to ChatGPT: The Evolution of Chatbots in Public Health. **XRDS: Crossroads, The ACM Magazine for Students**, v. 29, n. 3, p. 59–59, mar. 2023. Disponível em: [From ELIZA to ChatGPT: The Evolution of Chatbots in Public Health | XRDS: Crossroads, The ACM Magazine for Students](#). Acesso em: jun/2023.

HENRICKSON, L.; MEROÑO-PEÑUELA, A. Prompting Meaning: a Hermeneutic Approach to Optimising Prompt Engineering with ChatGPT. **AI & SOCIETY**, 4 set. 2023.

HOBERT, S.; FØLSTAD, A.; LAW, E.. Chatbots for Active Learning: A Case of Phishing Email Identification. **International Journal of Human-Computer Studies**, v. 179, p. 103108, 19 jul. 2023.

HOBERT, S.; MEYER, R. Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents. **Wirtschaftsinformatik und Angewandte Informatik**, p. 301–314, 1 jan. 2019.

HOLMES, W.; BIALIK, M; FADEL, C. **Artificial intelligence in education: promises and implications for teaching and learning**. Boston, Ma: The Center For Curriculum Redesign, 2019.

HWANG, G.-J.; CHANG, C.-Y. A Review of Opportunities And Challenges of Chatbots in Education. **Interactive Learning Environments**, p. 1–14, 18 jul. 2021.

IBM. **What is Artificial Intelligence (AI)?** Disponível em: <https://www.ibm.com/topics/artificial-intelligence>. Acesso em: Jun/2023.

IOWA STATE UNIVERSITY. Revised Bloom's Taxonomy. Disponível em: <https://www.celt.iastate.edu/instructional-strategies/effective-teaching-practices/revised-blooms-taxonomy/>. Acesso em: 5 dez. 2023.

JAKESCH, M. et al. **Co-Writing with Opinionated Language Models Affects Users' Views**. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). New York, NY, United States: Association for Computing Machinery, abr. 2023.

JALIL, S. et al. **ChatGPT and Software Testing Education: Promises & Perils**. IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW). 1 abr. 2023. Disponível em:
<<https://ieeexplore.ieee.org/abstract/document/10132255>>

JEON, J.; LEE, S. Large Language Models in Education: a Focus on the Complementary Relationship between Human Teachers and ChatGPT. **Education and Information Technologies**, v. 28, 2 maio 2023.

KAUFMAN, D. **Desmistificando a Inteligência Artificial**. Belo Horizonte: Autentica, 2022.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing systematic literature reviews in software engineering**. Keele University and Durham University Joint Report. 2004. Disponível em:
https://www.elsevier.com/_data/promis_misc/525444systematicreviewsguide.pdf. Acesso em: jun/2023.

KOSTKA, I.; TONCELLI, R. Exploring Applications of ChatGPT to English Language Teaching: Opportunities, Challenges, and Recommendations. **TESL-EJ**, v. 27, n. 3, 1 nov. 2023.

KUHAIL, M. A. et al. Interacting with Educational Chatbots: A Systematic Review. **Education and Information Technologies**, 9 jul. 2022.

LAAL, M. Lifelong Learning and Technology. **Procedia - Social and Behavioral Sciences**, v. 83, p. 980–984, jul. 2013. Disponível em: [Lifelong Learning and Technology - ScienceDirect](#). Acesso em: jun/2023.

LEHMAN, J. et al. **Evolution through Large Models**. 2022. Disponível em:
<https://arxiv.org/pdf/2206.08896.pdf>. Acesso em: jun/2023.

LUCKIN et al. **Intelligence unleashed: an argument for AI in education**. London: Pearson, 2016.

LUO et al. Aladdin's Genie or Pandora's Box for Early Childhood Education? Experts Chat on the Roles, Challenges, and Developments of ChatGPT. **Early Education and Development**, p. 1–18, 15 maio 2023. DOI: <https://doi.org/10.1080/10409289.2023.2214181>.

MARTHA, A. S. D.; SANTOSO, H. The Design and Impact of the Pedagogical Agent: A Systematic Literature Review. **The Journal of Educators Online**, v. 16, n. 1, jan. 2019.

MCCARTHY, J. **What is Artificial Intelligence?** 2007. Disponível em: <https://www-formal.stanford.edu/jmc/whatisai.pdf>. Acesso em: jun/2023.

MURGIA, E. et al. **Children on ChatGPT Readability in an Educational Context: Myth or Opportunity?** ACM Conference on User Modeling, Adaptation and Personalization. UMAP '23 ADJUNCT. New York, NY, United States: Association for Computing Machinery, 16 jun. 2023.

OKONKWO, C. W.; ADE-IBIJOLA, A. Chatbots Applications in Education: A Systematic Review. **Computers and Education: Artificial Intelligence**, v. 2, p. 100033, 2021.

OZMEN, G. et al. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. **International Journal of Human-Computer Interaction**, v. 39, n. 3, p. 391–437, jan. 2023. DOI: <https://doi.org/10.1080/10447318.2022.2153320>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/10447318.2022.2153320?scroll=top&needAccess=true&role=tab>. Acesso em: abril/2023.

PACK, A; MALONEY, J. Potential Affordances of Generative AI in Language Education: Demonstrations and an Evaluative Framework. **Teaching English with Technology**, v. 23, n. 2, 1 jan. 2023.

PASSMORE, J.; TEE, D. The Library of Babel: Assessing the Powers of Artificial Intelligence in Knowledge synthesis, Learning and Development and Coaching. **Journal of work-applied management**, 28 ago. 2023.

PÉREZ, J. Q.; DARADOUMIS, T.; PUIG, J. M. M. Rediscovering the Use of Chatbots in education: a Systematic Literature Review. **Computer Applications in Engineering Education**, v. 28, n. 6, 3 set. 2020.

PÉREZ-MARÍN, D. A Review of the Practical Applications of Pedagogic Conversational Agents to Be Used in School and University Classrooms. **Digital**, v. 1, n. 1, p. 18–33, 29 jan. 2021.

RADENSKY, M. et al. **“I Think You Might Like This”**: Exploring Effects of Confidence Signal Patterns on Trust in and Reliance on Conversational Recommender Systems. FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. New York, NY, United States: Association for Computing Machinery, 12 jun. 2023.

RAMANDANIS, D.; XINO GALOS, S. Designing a Chatbot for Contemporary Education: A Systematic Literature Review. **Information**, v. 14, n. 9, p. 503, 1 set. 2023.

RODA, C.; ANGEHRN, A.; NABETH, T. **Conversational agents for advanced learning: applications and research**. INSEAD - Centre for Advanced Learning Technologies. 2021. Fontainebleau, France. Disponível em:

https://www.researchgate.net/publication/228544951_Conversational_agents_for_advanced_learning_applications_and_research. Acesso em: jun/2023.

RUDOLPH, J.; TAN, Samson; TAN, Shannon. ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher education? **Journal of Applied Learning & Teaching**, v. 6, n. 1, 25 jan. 2023a.

RUDOLPH, J.; TAN, Samson; TAN, Shannon. War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The New AI Gold Rush and Its Impact on Higher Education. **Journal of Applied Learning & Teaching**, v. 6, n. 1, 25 abr. 2023b.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence**: A Modern approach. 4. ed. New Jearsey: Pearson, 2010.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of Research and Development**, v. 44, n. 1.2, p. 206–226, jan. 2000. Disponível em: [Some studies in machine learning using the game of checkers | IBM Journals & Magazine | IEEE Xplore](#). Acesso em: jun/2023.

SANTAELLA, L. **A Inteligência Artificial é Inteligente?** São Paulo: Grupo Almedina, 2023.

SCHLIMBACH et al. **A Literature Review on Pedagogical Conversational Agent Adaptation**. In: Conference: 26th Pacific Asia Conference on Information Systems (PACIS 2022). 2022. Disponível em:

https://www.researchgate.net/publication/361228223_A_Literature_Review_on_Pedagogical_Conversational_Agent_Adaptation. Acesso em: jun/2023

SCHMITT, A.; WAMBSGANSS, T.; LEIMEISTER, J. M. Conversational Agents for Information Retrieval in the Education Domain. **Proceedings of the ACM on Human-Computer Interaction**, v. 6, n. CSCW2, p. 1–22, 7 nov. 2022. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3555587>. Acesso em: jun/2023.

SHAWAR, B. A.; ATWELL, E. Chatbots: Are they Really Useful?. **Journal for Language Technology and Computational Linguistics**, [S. l.], v. 22, n. 1, p. 29–49, 2007. DOI: 10.21248/jlcl.22.2007.88. Disponível em: <https://jlcl.org/article/view/88>. Acesso em: abril/2023.

SMITH, A. et al. Old dog, new tricks? Exploring the potential functionalities of ChatGPT in supporting educational methods in social psychiatry. **International Journal of Social Psychiatry**, v. 69, n. 8, 30 jun. 2023.

SU, Y.; LIN, Y.; LAI, C. Collaborating with ChatGPT in argumentative writing classrooms. **Assessing Writing**, v. 57, p. 100752–100752, 1 jul. 2023.

TAMKIN et al. **Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models**. 2021. Disponível em: <https://arxiv.org/abs/2102.02503>. Acesso em: jun/2023

TLILI, A. et al. What If the Devil Is My Guardian angel: ChatGPT as a Case Study of Using Chatbots in Education. **Smart Learning Environments**, v. 10, n. 1, 22 fev. 2023.

VALMEEKAM et al. **Large Language Models Still Can't Plan** (A Benchmark for LLMs on Planning and Reasoning about Change). 2023. Disponível em: <https://arxiv.org/pdf/2206.10498.pdf>. Acesso em: jun/2023.

VASCONCELOS, M. A.; SANTOS, R. Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study. **Eurasia Journal of Mathematics, Science and Technology Education**, v. 19, n. 7, p. em2296, 1 jul. 2023.

VASWANI, A. et al. **Attention Is All You Need**. Disponível em: <<https://arxiv.org/abs/1706.03762>>.

WEBB, A. **The Big Nine**: how the tech titans and their thinking machines could warp humanity. New York: Public Affairs, 2020.

WIESER, M. et al. Investigating the Role of ChatGPT in Supporting Text-Based Programming Education for Students and Teachers. **Lecture Notes in Computer Science**, v. 14296, p. 40–53, 1 jan. 2023.

WINKLER, R.; SOELLNER, M. Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. **Academy of Management Proceedings**, v. 2018, n. 1, p. 15903, ago. 2018.

WOLLNY, S. et al. Are We There Yet? - A Systematic Literature Review on Chatbots in Education. **Frontiers in Artificial Intelligence**, v. 4, 15 jul. 2021.

YANG, Q. **Profiling Artificial Intelligence as a Material for User Experience Design**. 2020. 126 f. Tese (PhD) – Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh. Disponível em: https://kilthub.cmu.edu/articles/thesis/Profiling_Artificial_Intelligence_as_a_Material_for_User_Experience_Design/14376731. Acesso em: abril/2023.

ZAWACKI-RICHTER, O. et al. Systematic review of research on artificial intelligence applications in higher education – where are the educators? **International Journal of Educational Technology in Higher Education**, v. 16, n. 39, 28 out. 2019. Disponível em: <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-019-0171-0?fbclid=IwAR0vSk4s9y0V0vExpcAel6yL4LEb-PrNDnlreOB5WrGxlu8-3awpYGgK6Ig#citeas>. Acesso em: jun/2023.

ZHAI, X. et al. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020.

Complexity, p. 1–18, 20 abr. 2021. Disponível em:

<https://downloads.hindawi.com/journals/complexity/2021/8812542.pdf>. Acesso em: jun/2023.

12. APÊNDICE

Tabela 3 - Lista de Artigos Selecionados

Bases de dados	Título do Artigo	Autores e Ano de Publicação
SpringerLink	A Comprehensive AI Policy Education Framework for University Teaching and Learning	Chan (2023)
Scopus	Aladdin's Genie or Pandora's Box for Early Childhood Education? Experts Chat on the Roles, Challenges, and Developments of ChatGPT	Luo et al. (2023)
SpringerLink	Are We Dupes? Limitations of AI Systems: What Should Educators Do with Them?	Frick (2023)
Scopus	Chatbots for Active Learning: A Case of Phishing Email Identification	Hobert, Følstad, Law (2023)
Scopus	ChatGPT and Software Testing Education: Promises & Perils	Jalil et al. (2023)
Scopus	ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher education?	Rudolph, Tan, Tan (2023a)
Scopus	ChatGPT and the Frustrated Socrates	Gregorcic, Pendrill (2023)
Scopus	Children on ChatGPT Readability in an Educational Context: Myth or Opportunity?	
Scopus	Collaborating with ChatGPT in argumentative writing classrooms	Su, Lin, Lai (2023)
Scopus	Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study	Vasconcelos, Santos (2023)
SpringerLink	Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence	Cooper (2023)
Scopus	Exploring Applications of ChatGPT to English Language Teaching: Opportunities, Challenges, and Recommendations	Kostka, Toncelli (2023)
SpringerLink	GPT-3-Driven Pedagogical Agents to Train Children's Curious Question-Asking Skills	Abdelghani (2023)
SpringerLink	Investigating the Role of ChatGPT in Supporting Text-Based Programming Education for Students and Teachers	Wieser et al. (2023)
Web of Science (WoS)	Large Language Models in Education: a Focus on the Complementary Relationship between Human	Jeon, Lee (2023)

	Teachers and ChatGPT	
Scopus	Old dog, new tricks? Exploring the potential functionalities of ChatGPT in supporting educational methods in social psychiatry	Smith et al. (2023)
Scopus	Potential Affordances of Generative AI in Language Education: Demonstrations and an Evaluative Framework	Pack, Maloney (2023)
Scopus	The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age	Cardon et al. (2023)
Scopus	The Library of Babel: Assessing the Powers of Artificial Intelligence in Knowledge synthesis, Learning and Development and Coaching	Passmore, Tee (2023)
Scopus	Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT — a Case Study	Chaudhry et al. (2023)
Scopus	War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The New AI Gold Rush and Its Impact on Higher Education	Rudolph, Tan, Tan (2023b)
Scopus	What If the Devil Is My Guardian angel: ChatGPT as a Case Study of Using Chatbots in Education	Tlili et al. (2023)
ACM Digital Library	Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI Explanations.	Danry et al. (2023)
SpringerLink	Prompting Meaning: a Hermeneutic Approach to Optimising Prompt Engineering with ChatGPT	Henrickson, Meroño-Peñuela (2023)
Scopus	How General-Purpose Is a Language Model? Usefulness and Safety with Human Prompts in the Wild.	Casares et al. (2022)
IEEEExplore	Prompting Large Language Models with the Socratic Method	Chang (2023)
ACM Digital Library	Co-Writing with Opinionated Language Models Affects Users' Views	Jakesch et al. (2023)
