

# **Uso de *chatbots* na produção de resumos científicos: o que os resumos nos dizem da investigação? – Uma revisão sistemática**

**Helena Leitão**

0000-0002-8417-6657

Universidade Católica Portuguesa – Centro Regional do Porto, Portugal

hleitao@ucp.pt

**RESUMO:** Os resumos científicos são versões abreviadas, concisas e fiéis do conteúdo intelectual de um documento original. Eles podem ter uma estrutura explícita ou implícita e desempenham funções de alerta, antecipação, substituição e seleção de informação. Tradicionalmente, são produzidos por operadores humanos, mas, com o avanço da tecnologia, sistemas automáticos também passaram a ser utilizados, especialmente a partir da segunda metade do século XX. Os sistemas automáticos, inicialmente baseados na frequência e posição das palavras, evoluíram com a inteligência artificial. *Chatbots*, como o *Chat GPT*, lançado em 2022, revolucionaram a produção de texto, inclusive de resumos científicos, ao gerar conteúdo realista com base em grandes volumes de texto pré-existente. No entanto, há preocupações éticas sobre a autoria e a precisão científica desses resumos, uma vez que IA não pode ser responsável pelo conteúdo gerado.

Este estudo objetiva, através de resumos de artigos, analisar a utilização de ferramentas de IA na construção de resumos científicos, avaliando temas, contextos, métodos e conclusões de estudos recentes.

Metodologicamente, procedeu-se a uma revisão sistemática da literatura, utilizando-se as bases *Web of Science*, *Scopus* e *Ebsco Discovery Service*, focando em artigos dos últimos três anos.

A quantidade de estudos que existem sobre o tema, e que cujos resumos foram objeto de análise, demonstram que ainda há muita necessidade de

estudo e reflexão acerca deste tema, Destaca-se a dificuldade em distinguir resumos gerados por IA dos humanos e a controvérsia sobre a qualidade e ética do uso de IA. Apesar das melhorias na tecnologia, a validação humana ainda é crucial para garantir a precisão e a qualidade dos resumos, reforçando a importância do papel humano na supervisão e adaptação dos conteúdos gerados por IA.

**PALAVRAS-CHAVE:** Resumos científicos; Inteligência artificial; Resumos automáticos; Resumos manuais.

**ABSTRACT:** Scientific abstracts are abbreviated, concise, and faithful versions of the intellectual content of an original document. They can have an explicit or implicit structure and have functions such as alerting, anticipating, substituting, and selecting information. Traditionally, they are produced by human operators, but with technological advancements, automated systems have also been used, especially since the second half of the 20th century. Initially, these automated systems were based on word frequency and position, but they have evolved with artificial intelligence. *Chatbots*, like *Chat GPT*, launched in 2022, have revolutionized text production, including scientific abstracts, by generating realistic content based on large volumes of pre-existing text. However, there are ethical concerns about authorship and the scientific accuracy of these abstracts since AI cannot be held responsible for the generated content.

This study aims to analyse the use of AI tools in constructing scientific abstracts through article summaries, evaluating themes, contexts, methods, and conclusions of recent studies.

Methodologically, a systematic literature review was conducted using the Web of Science, Scopus, and Ebsco Discovery Service databases, focusing on articles from the last three years.

The existing studies, whose summaries were analyzed, demonstrate a significant need for further study and reflection on this topic. There is a noted difficulty in distinguishing AI-generated abstracts from human ones and controversy regarding the quality and ethics of using AI. Despite technological improvements, human validation remains crucial to ensure the accuracy and quality of abstracts, reinforcing the importance of human oversight and adaptation of AI-generated content.

**KEYWORDS:** Scientific abstracts; Artificial intelligence; Automatic abstracts; Human abstracts.

## Introdução

Um resumo científico consiste numa versão abreviada, concisa e fiel do conteúdo intelectual do documento que lhe deu origem (Bernier, 1968; Lancaster & Neway, 1982; Leitão et al., 2015; Maizell et al., 1979; Moore, 2009; Pinto Molina, 2001; Simões, 2015).

Tendo como funções alertar, antecipar, substituir e selecionar informação (Simões, 2015), no que se refere à sua construção, apesar de tradicionalmente estar associada a operadores humanos, particularmente ao autor do documento primário, fruto de um trabalho intelectual (Moreiro Gonzalez, 2004; Pinto Molina, 2001), pode ser produzida por outros operadores humanos (especialista na matéria ou resumidor profissional), ou por sistemas automáticos.

Os sistemas automáticos de produção de resumos surgiram na segunda metade do século XX (Cunha et al., 2009; Gillick, 2011) com Hans Peter Luhn (Kobashi, 1996). A técnica de construção de resumos automáticos até ao surgimento dos softwares atuais de inteligência artificial (IA), estava relacionada com uma lógica de significância das frases (Luhn, 1958).

Os *chatbots*, criados com o objetivo de compreenderem, interpretar e gerarem texto idêntico ao produzido pelo ser humano (Rossetini et al., 2023), têm atraído os investigadores e revolucionado a forma como a informação científica é produzida (Osama & Afridi, 2023), e vêm colmatar as lacunas que existiam nos processos automáticos de produção de resumos baseados em métodos extrativos, em que o produto final era uma junção de extratos do texto original.

O lançamento do *chatbot Chat GPT*, em 2022, revolucionou a forma como a informação é produzida (Hua et al., 2023), na medida em que gera texto realista em resposta a solicitações específicas do utilizador (Else, 2023). Constituindo-se como um *large language model* (LLM), baseia-se “on neural networks that learn to perform a task by digesting huge amounts of existing human-generated text” (p. 423).

Uma questão importante, e para a qual o *Committee on Publication Ethics* (COPE) salienta é a autoria da informação. Para o COPE (2023),

as ferramentas de IA, não podem ser indicadas como autores, já que “AI tools cannot meet the requirements for authorship as they cannot take responsibility for the submitted work”.

A verdade é que frequentemente são apresentadas vantagens e desvantagens estimadas do uso destes sistemas, havendo uma constante referência às alucinações dos sistemas de IA, que acabam por produzir texto que não é cientificamente válido (Else, 2023). Para Rossettini et al. (2023), a forma como direcionamos o *chatbot*, ou seja, as questões e as indicações que lhe damos, são determinantes para o produto final, sendo uma das suas *fraquezas* a dependência que tem das orientações humanas.

Vários estudos e discussões têm sido promovidas acerca da aplicabilidade da IA e dos *chatbots* na análise e produção de informação científica. Assim, questiona-se que investigação está a ser feita e que conclusões já se alcançaram acerca da sua utilização na produção de resumos científicos.

## Objetivo e Metodologia

O trabalho tem como objetivo concluir acerca dos temas e contextos dos estudos mais atuais no que se refere ao uso de ferramentas de IA – particularmente *chatbots* – na construção de resumos científicos. Esta aferição é relevante na medida em que nos permite compreender, por um lado, que pontos-chave estão a ser estudados, com que objetivos, que métodos estão a ser utilizados, que resultados têm sido alcançados e, principalmente, que conclusões podem ser retiradas; e por outro contribuir para a reflexão sobre o tema no sentido de desenvolvimento de estudos mais aprofundados, ou de reflexão acerca de tópicos que estão a ser marginalizados nas análises.

Do ponto de vista metodológico, optou-se por uma revisão sistemática da literatura, incidindo a análise nos resumos de artigos. Utilizaram-se as bases de dados referenciais Web Of Science e Scopus e a plataforma Ebsco Discovery Service. Com base nas palavras-chave em inglês *scientific abstract*, *artificial intelligence*, *chatbots* criou-se a expressão “*scientific abstract\**” AND (“*artificial intelligence*” OR IA OR *Chatbots*). A pesquisa incidiu nos campos de *título*, *resumo* e *palavras-chave* e optou-se por

inserir um limitador cronológico, restringindo-se a documentos publicados nos últimos 3 anos (2022 atualidade-junho de 2024).

Foram recuperados 46 artigos, sendo identificados 20 duplicados. Foi ainda removido um documento que sugeria ser um duplicado e dois documentos que não apresentavam resumos. Após triagem dos 12 resultados restantes através dos resumos, foram excluídos 5 resultados que não estavam relacionados efetivamente com o tema em análise. Assim, concluiu-se um total de 7 resumos de artigos para revisão.

**Quadro 1** – Estudos incluídos na revisão

| <b>Autor</b>                 | <b>Título</b>  |
|------------------------------|--|
| Kim H.J. et al.              | <i>Assessing the Reproducibility of the Structured Abstracts Generated by ChatGPT and Bard Compared to Human-Written Abstracts in the Field of Spine Surgery: Comparative Analysis</i> |
| Eppler M.B. et al.           | <i>Bridging the Gap Between Urological Research and Patient Understanding: The Role of Large Language Models in Automated Generation of Layperson's Summaries</i>                      |
| Howard F.M. et al.           | <i>Characterizing the Increase in Artificial Intelligence Content Detection in Oncology Scientific Abstracts From 2021 to 2023</i>   |
| Cheng S.-L. et al.           | <i>Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study</i>                             |
| Theocharopoulos P.C. et al.  | <i>Detection of Fake Generated Scientific Abstracts</i>  |
| Hong-Uyen Hua, et al.        | <i>Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots</i>   |
| Muhammad Osama; Sabah Afridi | <i>ChatGPT: A New Era in Research Writing Assistance</i>   |

No que se refere à análise dos resumos, do ponto de vista formal, analisou-se a sua autoria – se identificada como humana ou produzida por IA. Do ponto de vista do conteúdo, analisou-se a área do conhecimento geral do artigo, com vista a concluir se há uma tendência para a publicação

sobre este tema nalgum contexto científico em particular; os elementos estruturantes – contexto, procurando perceber o que desencadeou o estudo; os objetivos que nortearam os estudos; a metodologia seguida as conclusões retiradas dos estudos; e as palavras-chave, aferindo que termos são mais utilizados na representação destes assuntos.

Para a organização da informação recolhida, utilizou-se um quadro.

## Resultados

Após análise dos resumos dos artigos, verificou-se que, no que se refere à autoria, nenhum dos artigos identificava a responsabilidade pela construção do texto. Pelo facto de não haver uma referência ao uso de ferramentas de IA na sua construção, e pelas suas características textuais e estruturais, infere-se que os 7 resumos foram construídos por um operador humano, no caso o/os autores do artigo original.

Relativamente às áreas do conhecimento, 6 artigos enquadram-se na área da medicina tanto tematicamente quanto ao nível da revista onde foram publicados. O artigo restante, não identifica nenhuma área em concreto.

No que se refere ao enquadramento, especificamente ao pressupostos que potenciaram estes estudos: a) dificuldade na distinção entre resumos produzidos por IA e por humanos; b) controvérsia no uso da IA para a produção de conteúdo científico particularmente resumos; c) falta de estudos acerca das diferenças na capacidade dos vários softwares de IA; d) impacto significativo dos *chatbots*, tanto para a literacia em saúde como para a investigação; e) falta de estudos acerca das desvantagens dos *softwares* de IA na produção de resumos, particularmente da criação de conteúdo impreciso proveniente de alucinações; b) inevitabilidade da integração da IA no quotidiano da população.

No que se refere aos objetivos, podem ser resumidos em 5: a) comparação da similaridade entre o produto da IA e o operador humano; b) avaliação da probabilidade dos resumos terem sido produzidos por AI; c) comparação da qualidade do produto humano e de AI; d) Capacidade de o software de AI gerar resumos adequados ao público leigo; e) avaliação da aplicabilidade de um modelo de AI na construção de resumos;

Do ponto de vista da metodologia, verificou-se quatro métodos distintos relacionados com produção de resumos e com a avaliação do produto final: a) comparação entre resumos produzidos pela IA e pelo homem com base em instruções aos autores; b) utilização de validadores humanos, especialistas na matéria, para a deteção cega de resumos produzidos por IA e por operadores humanos; c) aplicação de métodos automáticos para a deteção de conteúdo produzido por IA; d) Utilização algoritmos para a verificação da similaridade entre os dois tipos de resumos.

A utilização de validadores humanos e de métodos automáticos para deteção de conteúdo produzido por IA, surgem em na maioria dos artigos.

Importa salientar que, do ponto de vista de *chatbots*, 6 artigos (em 7), analisaram o Chat GPT. Apenas dois artigos tiveram como base outros *chatbots* e um dos artigos não refere nenhum em concreto.

Relativamente às conclusões dos estudos, podemos sistematizá-las em 9 pontos-chave: a) na comparação entre o *Chat GPT* e o *Bard*, ambos podem ser usados para ajudar a produzir resumos, mas a maioria dos resumos são considerados antiéticos devido às altas taxas de plágio e deteção de IA; b) tendo em consideração as diretrizes de formatação, os resumos gerados pelo *Chat GPT*, em comparação com outros *softwares*, demonstram superioridade; c) o *ChatGPT* pode criar resumos precisos para o público leigo a partir de resumos científicos, com instruções bem elaboradas que melhoram a facilidade de utilização da informação; d) o uso do *ChatGPT* para criação de resumos científico pode não apresentar problemas de similaridade quando a sua fonte são textos completos reais escritos por humanos; e) tendo em consideração que os revisores humanos não distinguem com precisão os resumos produzidos por humanos e os produzidos por IA, é fundamental avaliar os limites éticos do uso de programas de IA; f) verifica-se uma presença crescente de conteúdo gerado por IA; g) conclui-se que o tamanho do resumo ou o tipo de artigo não afetam a probabilidade de o conteúdo ser gerado por IA; h) recomenda-se prudência no uso destes recursos de IA para fins de educação em saúde ou académicos; i) embora os resumos produzidos por IA sejam satisfatórios, é necessária uma verificação especializada para melhorar a precisão.

No que se refere às palavras-chave, 2 dos 7 resumos não apresentam nenhuma. Os termos presentes nos restantes 5 artigos, foram sistematizados num quadro. Salientamos a quantidade de formas de representação dos conceitos que foram a base da nossa pesquisa: 3 termos representativos do conceito de IA, 4 representativos do conceito de modelos de linguagem, um representativo dos *Chatbots* e 3 representativos dos resumos científicos. As variações estão relacionadas com abreviaturas, plurais e singulares e relações entre conceitos mais gerais e mais específicos.

### **Conclusão – O que os resumos nos dizem?**

O uso de ferramentas de IA está claramente disseminado na produção de informação científica. Ao nível dos estudos sobre a produção de resumos, conclui-se, em primeiro lugar, que o número de estudos publicados ainda são muito limitados.

Um aspeto importante, tem a ver com o facto de nenhum dos resumos se identificar como produzido por IA, inferindo-se que são produto do trabalho humano. Esta inferência justifica-se, tendo em consideração que seria essencial essa informação estar descrita na metodologia caso um sistema de IA tivesse sido utilizado, indo ao encontro do entendimento do COPE (2023) de que “Authors who use AI tools in the writing of a manuscript, production of images or graphical elements of the paper, or in the collection and analysis of data, must be transparent in disclosing in the Materials and Methods (or similar section) of the paper how the AI tool was used and which tool was used”.

Estes *resumos* também reforçam o facto de continuar a ser a área da saúde a pioneira na investigação acerca dos resumos científicos, produzindo quase todos os estudos inovadores. O artigo acerca da produção de resumos para o público leigo relembra-nos um aspeto fundamental – a literacia da população. Apesar de, historicamente, a literatura científica ter como foco a comunicação entre pares (Borges, 2006), a informação produzida deve estar à mercê do uso e discussão pública. No caso da saúde, por exemplo, “Medical journals, intended primarily for specialists, have widened their audience with the advent of the World Wide Web” (Splendiani et al., 2014, p. 419).

Constata-se o facto de que os sistemas de IA estão implantados e demonstram melhores resultados e mais alternativas a cada dia, mas o *Chat GPT* continua a ser o *chatbot* mais reconhecido.

Do ponto de vista metodológico, conclui-se que os estudos têm uma vertente tendencialmente comparadora dos dois tipos de resumos. Da mesma forma, utilizam validadores, sendo que os validadores humanos têm um papel muito relevante, particularmente numa fase em que a aplicabilidade de *chatbots* na construção de resumos apresenta taxas significativas de erro e/ou apresenta claramente características que fazem com que o conteúdo seja imediatamente identificado como produzido por IA.

Salienta-se a dependência que os *chatbots* têm da interação com o homem, e o quanto as orientações humanas são importantes para a precisão. No caso dos resumos, este aspeto é muito significativo, podendo ser determinante na escolha do sistema a utilizar, na medida em que permite que possamos indicar algumas especificidades necessárias, como as orientações das revistas, a estrutura pretendida, dimensões, entre outros aspetos. Assim, verifica-se que o operador humano continua a ter um papel central na garantia da qualidade dos processos automáticos – pela intervenção que tem através das *prompts* – e do produto final, através da sua adaptação, correção e validação. Com isto, o princípio da prudência deve ser sempre considerado, principalmente quando a aplicação destes sistemas afeta aspetos essenciais da vida humana e a responsabilidade final pelo produto de informação é sempre do autor humano que recorreu à ferramenta.

### Referências bibliográficas

- Bernier, C. (1968). Abstracts and abstracting. In *Encyclopedia of library and information science* (Vol. 1, pp. 16-38). Marcel Dekker Inc.
- Borges, M. M. (2006). *A Esfera: Comunicação Académica e Novos Media* [University of Coimbra]. <http://hdl.handle.net/10316/8557>
- COPE. (2023, February 13). *COPE position statement*. <https://publicationethics.org/cope-position-statements/ai-author>
- Cunha, I., Torres-Moreno, J. M., Velazquez-Morales, P., & Vivaldi, J. (2009). Un algoritmo lingüístico-estadístico para resumen automático de textos especializados. *Linguamática*, 2, 67-80. <http://www.linguamatica.com/index.php/linguamatica/article/view/33/0>

- Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, 613. <https://doi.org/10.1101/2022.12.23.521610>
- Gillick, D. J. (2011). The Elements of Automatic Summarization. *EECS UC Berkeley Thesis*, 53(9), 1689-1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Hua, H. U., Kaakour, A. H., Rachitskaya, A., Srivastava, S., Sharma, S., & Mammo, D. A. (2023). Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots. *JAMA Ophthalmology*, 141(9), 819-824. <https://doi.org/10.1001/jamaophthalmol.2023.3119>
- Kobashi, N. Y. (1996). Análise documentária e representação da informação. *Informare: Cadernos Do Programa de Pós-Graduação Em Ciência Da Informação*, 2(2), 5-27. [https://edisciplinas.usp.br/pluginfile.php/359680/mod\\_resource/content/1/Kobashi AD.pdf](https://edisciplinas.usp.br/pluginfile.php/359680/mod_resource/content/1/Kobashi AD.pdf)
- Lancaster, F. W., & Neway, J. M. (1982). The future of indexing and abstracting services. *Journal of the American Society for Information Science*, 33(3), 183-189. <https://doi.org/10.1002/asi.4630330315>
- Leitão, H., Borges, M. M., & Simões, M. da G. (2015). Resumos Científicos em Saúde – Estudo exploratório de resumos de artigos das revistas Nature e PLOS One. *Ciência e Sociedad*. <https://doi.org/10.13140/RG.2.1.3056.2641>
- Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159-165. <https://doi.org/10.1147/rd.22.0159>
- Maizell, R., Smith, J., & Singer, T. (1979). *Abstracting scientific and technical literature : an introductory guide and text for scientists, abstractors, and management*. R. E. Krieger Pub. Co.
- Moore, A. (2009). What's in an abstract? *BioEssays*, 31(3), 261. <https://doi.org/10.1002/bies.200900023>
- Moreiro Gonzalez, J. (2004). *El contenido de los documentos textuales: su analisis y representación mediante el lenguaje natural*. Ediciones Trea.
- Osama, M., & Afridi, S. (2023). ChatGPT: A New Era in Research Writing Assistance. In *Journal of the Pakistan Medical Association* (Vol. 73, Issue 9, pp. 1929-1930). Pakistan Medical Association. <https://doi.org/10.47391/JPMA.9183>
- Pinto Molina, M. (2001). *El resumen documental: Paradigmas, modelos y métodos*. Fundación Germán Sánchez Ruipérez.
- Rossettini, G., Cook, C., Palese, A., Pillastrini, P., & Turolla, A. (2023). Pros and Cons of Using Artificial Intelligence Chatbots for Musculoskeletal Rehabilitation Management. In *Journal of Orthopaedic and Sports Physical Therapy* (Vol. 53, Issue 12, pp. 728-734). Movement Science Media. <https://doi.org/10.2519/jospt.2023.12000>
- Simões, M. da G. (2015). *Resumo Documental: um incursão à (des)construção concetual na literatura científica*. Imprensa da Universidade de Coimbra.