# UTILIZAÇÃO DE IA GENERATIVA PARA AUTOMATIZAR O PROCESSO DE FILTRAGEM DE ARTIGOS DURANTE O DESENVOLVIMENTO DE UMA REVISÃO SISTEMÁTICA DA LITERATURA.

**Cainã Antunes Silva[1], Antônio Cesar Germâno Martins[2]**

*[1]Universidade Estadual Paulista (UNESP), Sorocaba, Brasil (caina.antunes@unesp.br)*
*[2] Universidade Estadual Paulista (UNESP), Sorocaba, Brasil*

*Resumo:* A revisão de literatura, embora essencial, é uma tarefa que exige muita dedicação do pesquisador. Com o advento da Inteligência Artificial (IA) generativa, questiona-se o seu emprego como ferramenta de auxílio ao pesquisador durante o desenvolvimento desta etapa. Este trabalho utiliza o *chatGPT* para verificar a similaridade entre títulos de artigos, ajudando o pesquisador a selecionar, de forma automatizada, os trabalhos mais relevantes que servirão como base para seu projeto.

*Palavras-chave*: ChatGPT, Inteligência Artificial, Revisão de Literatura.

## INTRODUÇÃO

O conhecimento científico, produto do método científico, é uma conquista recente da humanidade. Precedida pelo conhecimento filosófico e empírico, a ciência moderna parte da determinação de um objeto de investigação e com o método que fará o controle desse conhecimento. A abordagem que a ciência faz da realidade permite a previsibilidade dos fenômenos, o que, consequentemente possibilita um maior poder para a transformação da natureza. Desta característica da ciência, resultou a tecnologia, que transformou o estilo de vida humano, sobretudo a partir do século XX. (Rampazzo, 2002)

O método científico, é composto por uma série de etapas, dentre elas está a revisão de literatura que, segundo Brizola e Fantin (2016), auxilia na delimitação do problema da pesquisa, na busca por novas linhas de investigação do problema, evitando abordagens infrutíferas ao identificar trabalhos já realizados.

A revisão de literatura é uma etapa indispensável para o desenvolvimento da metodologia científica, entretanto, é uma tarefa que exige muito tempo de dedicação do pesquisador.

Com o advento da Inteligência Artificial (IA) generativa, e sua ampla utilização em diversas áreas, formulou-se como hipótese que subsidiará este estudo, a empregabilidade desta tecnologia emergente como ferramenta de auxílio ao pesquisador durante o desenvolvimento da revisão de literatura.

Devido a sua capacidade cognitiva de interpretação de textos, este estudo irá apresentar como metodologia, o emprego de IA generativa para verificação de similaridade entre títulos de artigos, ajudando o pesquisador a classificar de forma mais automatizada, os trabalhos mais relevantes que lhe auxiliaram na escrita de sua revisão de literatura.

## MATERIAL E MÉTODOS

O método científico se dedica a criar conhecimento científico. Segundo Marconi e Lakatos (2005) trata-se das atividades sistemáticas e racionais que objetivam o alcance de conhecimentos válidos e verdadeiros, detectando erros e auxiliando as decisões dos pesquisadores.

O método científico tem início com a identificação de uma problemática seguida de sua colocação precisa. Posteriormente é realizada uma pesquisa de conhecimentos ou instrumentos relevantes ao problema, destaca-se esta etapa que, por ser o objeto de experimentação deste artigo, será retomada adiante. Utilizando os resultados obtidos na etapa anterior, tenta-se solucionar o problema proposto. Em caso de insucesso formula-se novas ideias que possam gerar uma solução. Por fim deve-se investigar as consequências da solução encontrada e colocá-la à prova. (Miguel et al., 2022)

A revisão da literatura é essencial para o desenvolvimento do método científico, auxiliando o pesquisador, segundo Brizola e Fantin (2016), na delimitação do problema da pesquisa, na busca por novas linhas de investigação do problema, a evitar abordagens infrutíferas, identificar trabalhos já realizados possibilitando a adoção de novas abordagem e assim evitando que o pesquisador faça o que já foi feito, diga o que já foi dito, tornando seu trabalho irrelevante.

De Souza et al. (2018) descreve 14 tipos de revisões da literatura em sua análise. Dentre elas, está a revisão sistemática, escolhida para o desenvolvimento deste artigo.

Compõe as etapas de uma revisão sistemática da literatura, a escolha das fontes de buscas da temática, definição de uma estratégia de busca para o viés da pesquisa, avaliação dos estudos selecionados, escolha das ferramentas utilizadas na síntese dos resultados e apresentação do estudo. (Brizola e Fantin, 2016)

Como visto anteriormente, a revisão da literatura é uma técnica fundamental para reunir conhecimentos e instrumentos já validados pela comunidade científica e que possam ser úteis para a solução do problema proposto por estudos que sigam a metodologia científica. Entretanto, mediante a quantidade massiva e crescente de trabalhos acadêmicos existentes, desenvolver uma revisão sistemática da literatura tem sido um grande desafio, sobretudo devido ao grande volume de trabalhos disponíveis em repositórios digitais e o tempo necessário para selecionar e avaliar os materiais mais relevantes para uma pesquisa.

Sendo assim, a colocação do problema central, ao qual se baseia este artigo, é: "Como automatizar as etapas de busca e avaliação de materiais acelerando o desenvolvimento de revisões sistemáticas da literatura?"

A hipótese proposta é: "Dado o funcionamento de IAs generativas como o *chatGPT*, esta tecnologia emergente teria capacidade de realizar tal automatização? Se sim, com qual confiabilidade?"

Atualmente, há uma crescente popularização de chats que utilizam IA para interagir com o público, e tornou-se comum observar o uso desta tecnologia para o auxílio no desenvolvimento de trabalhos acadêmicos.

A Microsoft (2024), apresenta em sua plataforma de treinamentos os conceitos básicos de IA generativa, os quais serão apresentados nos próximos parágrafos.

Os aplicativos de IA generativos atuais utilizam tipos especializados de aprendizagem de máquina denominados *Large Language Models (LLM)*, ou em português, Modelos de Linguagem Grandes. Estes modelos permitem a execução de tarefas de *Natural Language Processing (NLP)* que incluem: classificação do texto em linguagem natural, resumo de textos, geração de nova linguagem natural e, com maior relevância para este artigo, comparar várias fontes de texto quanto à similaridade semântica.

Ao longo dos anos, os modelos de processamento de linguagem natural evoluíram substancialmente, sendo que os *LLMs* atuais, como os utilizados pelo *chatGPT*, são baseados na arquitetura de transformadores, treinados com grandes volumes de textos, permitindo a representação semântica entre palavras e consequentemente a elaboração de sequências prováveis de texto que façam sentido. Esta arquitetura é composta por um bloco codificador, que cria representações semânticas do vocábulo de treino, e uma bloco decodificador que gera novas sequencias prováveis de texto.

Inteligências Artificiais generativas construídas sobre a arquitetura de transformadores tem seu treinamento baseado em uma técnica de extração e indexação de *tokens*, onde cada *tokens* representa uma palavra ou palavra parcial ou conjunto de palavra e pontuação.

A simples *"tokenização"* não confere ao modelo a capacidade de gerar relações semânticas, para isso são feitas inserções de vetores numéricos a cada *token* de forma que se representados graficamente cada *token* ocuparia um lugar em um espaço multidimensional, sendo a distância entre cada palavra uma medida de sua similaridade. Em outras palavras o treinamento de *LLMs* consiste na extração, e inserções de *tokens* que posteriormente podem ser relacionados em forma de *clusters* conferindo-lhes dessa forma relações semânticas.

Este mecanismo se encaixa perfeitamente para solução do problema colocado para este artigo, a verificação da similaridade entre os títulos de estudos obtidos pela busca em repositórios acadêmicos digitais durante o desenvolvimento de uma revisão sistemática da literatura.

Para testar esta hipótese, será utilizada a estratégia de filtros proposto por De Lima Andrade et al. (2021):

1. Definição de descritores baseados no problema para busca em diferentes bases de dados digitais.
2. Exclusão de artigos duplicados devido a busca em diferentes bases de dados digitais.
3. Exclusão de artigos baseado em títulos fora do contexto da problemática proposta.
4. Exclusão baseada em resumos sem elementos que satisfaçam a problemática proposta para a presente revisão.
5. Exclusão baseada no texto completo de artigos que não contemplem os elementos do objetivo da revisão.

A hipótese proposta neste artigo atua na facilitação da execução das etapas 1, 2 e 3 da estratégia supracitada podendo ser estendida também para a etapa 4.

As próximas seções se dedicam ao desenvolvimento, testes e avaliação da hipótese aplicada a elaboração da revisão sistemática da literatura para um trabalho onde os autores, utilizando IA visam melhorar imagens capturadas em ambientes de baixa luminosidade. Este trabalho foi selecionado para

testar a hipótese proposta neste artigo, mas vale ressaltar que a metodologia apresentada, pode ser aplicada ao desenvolvimento de revisões sistemática da literatura em geral sem restrições de áreas.

**1. Primeiro filtro: Busca por descritores em base de dados digitais**

Seguindo a estratégia de filtros sequenciais para o desenvolvimento de revisões da literatura discorrido no item 2.3, a primeira etapa deste processo foi a escolha dos repositórios digitais, sendo eleitos 4 deles amplamente utilizados pela comunidade acadêmica: *Google Scholar, ScienceDirect, Scopus e Web of Science.*

Na sequência foram definidos o descritores a serem utilizados para a busca nos repositórios escolhidos. Para melhor compreensão da escolha destes descritores faz se necessário conhecer o contexto ao qual eles foram aplicados.

Sendo assim, o problema levantado durante o desenvolvimento do trabalho foi: Como melhorar a qualidade de fotografias capturadas em ambientes com pouca iluminação? Seguida do questionamento: "Um modelo de Inteligência Artificial seria capaz de atingir resultados satisfatórios no processo de melhoria dessas imagens?"

Dada a contextualização, foram definidos descritores que pudessem representar a ideia central por traz do problema e hipótese propostos. Estes descritores foram agrupados em três conjuntos, onde cada conjunto possui descritores variantes, mas de mesmo sentido semântico, e os três conjuntos combinados corroboram a ideia central desejada.

1. "*Low light*", "*dark*", "*low lighting*" e "*low luminance*".
2. "*Photography*" e "*image*"
3. "*Intelligent processing*", "*artificial intelligence*", "*deep learning*" e "*machine learning*".

Foram considerados apenas materiais escritos na língua inglesa na busca e os conjuntos de descritores foram aplicados sobre a busca por título, o que significa que trabalhos com títulos que não atendiam as regras dos descritores foram desconsiderados. Os conjuntos de descritores foram combinados utilizando-se operadores lógicos como mostrado a seguir:

("*low light*" ∨ "*low lighting*" ∨ "*low luminance*" ∨ "*dark*") ∧ ("*photography*" ∨ "*image*") ∧ ("*intelligent processing*" ∨ "*deep learning*" ∨ "*artificial intelligence*" ∨ "*machine learning*")

Os resultados de busca em cada repositório foram exportados no formato de arquivo de banco de dados *BibTeX*, que é formado por uma lista de entradas, sendo cada entrada um item bibliográfico. A escolha deste formato de arquivo foi feita para facilitar a execução dos próximos passos descritos nesta seção.

**2. Segundo filtro: exclusão de artigos duplicados**

É comum que um mesmo artigo seja citado em diferente bases de dados de trabalhos acadêmicos, e por isso, ao realizar buscas em diferentes repositórios, pode-se obter arquivos duplicados.

Esta seção se dedica a descrever o processo de exclusão de arquivos duplicados obtidos na busca da etapa anterior.

Para isso, os quatro arquivos de banco de dados *BibTeX* obtidos de cada repositório foram compilados em um único arquivo. Então, cada item bibliográfico da lista de entradas do arquivo *BibTeX* compilado foi verificado e os títulos duplicados excluídos.

A fim de automatizar esse processo de verificação, a manipulação dos arquivos foi realizada em ambiente de *"notebook"* do *Google Colaboratory* utilizando-se a biblioteca *bibtexparser* da linguagem de programação *Python*.

A biblioteca *bibtexparser* possui um método denominado *.load("file.bib")* que converte o arquivo de banco de dados *BibTeX* em uma lista de objetos *Python*. Estes objetos possuem métodos próprios e por estarem dentro de uma lista, são iteráveis.

Utilizando um *loop foreach* é possível iterar sobre cada entrada do arquivo BibTeX convertido em objeto. A cada iteração é possível se obter o título de cada item bibliográfico através do método *.get("key")* passando como parâmetro a chave "*title*".

Para fins de normalização, a cada título recuperado, foi aplicado o método *str.lower("string")* que faz com que todos os caracteres que o compõe assumam o formato minúsculo. Além disso foi aplicado o método *.sub("regular expression", "substitute string", "original string")* que através da expressão regular [^\w\s] que nega caracteres alfanuméricos (\w) e espaços (\s), substituiu todo tipo de caractere especial por uma uma *string* vazia. Esta normalização dos títulos foi essencial para a comparação direta entre *strings* durante a verificação de títulos repetidos.

A verificação de títulos duplicados se deu a partir da seguinte lógica: foi criada uma lista vazia de objetos, a cada iteração foi verificado se o título normalizado já existia na lista, em caso afirmativo a entrada do arquivo *BibTeX* é descartada, caso contrário é adicionada à lista, desta forma garantiu-se que cada entrada fosse adicionada apenas uma vez à lista eliminando entradas duplicadas.

Ao final do *loop* os objetos salvos na lista foram convertidos de volta para um arquivo de banco de dados *BibTeX*.

**3. Terceiro filtro: exclusão por relevância do título**

Nesta etapa da revisão sistemática da literatura, realizada uma nova filtragem do material obtido na etapa anterior com base na relevância de seus títulos com relação a colocação do problema e hipótese tratados em seu estudo. Desta forma, títulos distantes da temática do trabalho podem ser descartados.

A *Application Programming Interface (API)* do *chatGPT* foi a ferramenta escolhida para a execução desta tarefa. No *site* da *Openai* foi criada uma conta gratuita, limitada a cem requisições diárias no período aproximado de três meses, e então gerada uma chave de API.

No *Google Colab* foi instalada a biblioteca *openai* e configurada a conexão com a API através da chave criada anteriormente

A solução proposta envolve solicitar à IA generativa que compare a similaridade entre um texto padrão elaborado pelos autores para representar um título relevante ao artigo em desenvolvimento com os títulos provenientes da busca nos repositórios.

O texto padrão foi: "*intelligent processing for low-light image enhancement*".

A IA generativa deve comparar semanticamente o texto padrão com cada título fornecido pela etapa anterior, devolvendo como resposta um valor numérico que represente o grau de similaridade semântica entre as duas sentenças, sendo 0 nada similar e 1 totalmente similar.

A solicitação completa enviada para a IA generativa foi a seguinte:

*"Check the similarity between " + comparison_sentece + " and " + title + ", answer with a number between 0 and 1, where 0 means completely not similar and 1 means completely similar. The answer must contain only the number without texts".*

Sendo que as variáveis de concatenação *camparison_sentece* e *title* são respectivamente o texto padrão e o título extraído do arquivo de banco de dados *BibTeX* gerado no segundo filtro.

Desta vez foram criadas duas listas de objetos vazias uma para títulos relevantes e outra para títulos rejeitados. Cada resposta recebida do *chatGPT* foi classificada com base em um limiar, sendo que os títulos apontados pela IA com similaridade maior ou igual 0.55 foram incluídos a lista de títulos relevantes enquanto títulos com similaridade inferior a este limiar foram inseridos na lista de títulos rejeitados.

Ao fim desse processo cada lista foi convertida e salva no formato *BibTeX*.

Com o intuito de validar a hipótese, os mesmo títulos classificados anteriormente foram reclassificados em relevantes ou rejeitados manualmente pelos autores.

Todos os resultados obtidos foram carregados em um *dataframe* contendo três colunas, a primeira com o título de cada artigo, a segunda e a terceira com valores binários representando respectivamente as classificações da IA generativa e as classificações manuais dos autores, nestas duas últimas colunas valores 0 representam títulos rejeitados e 1 os títulos relevantes. Estes dados serão analisados adiante na seção resultados.

## RESULTADOS E DISCUSSÃO

Como pode ser observado na Figura 1 o primeiro filtro resultou em 170 artigos provenientes dos repositórios escolhidos ao se aplicar os descritores definidos, sendo 39 artigos do repositório *Web of Science*, 39 artigos do repositório *Scopus*, 26 artigos do repositório *ScienceDirect* e 66 artigos do repositório *Google Scholar*.

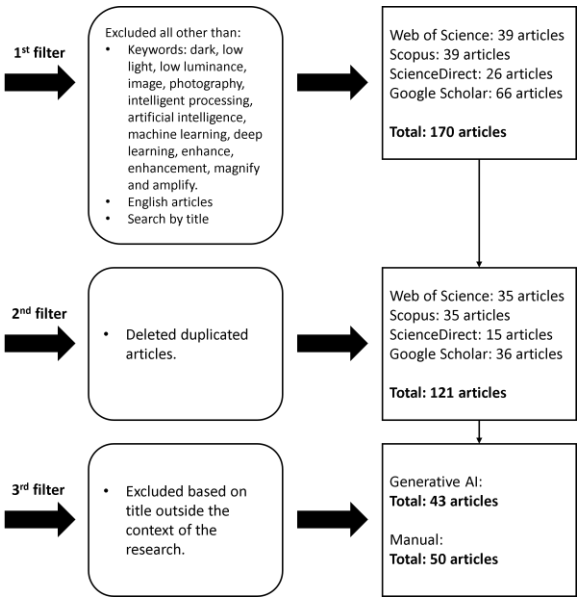


Figura 1. Resultado dos filtros.

Já o segundo filtro eliminou 49 artigos duplicados, restando 121 trabalhos distribuídos da seguinte forma: 35 artigos do repositório *Web of Science*, 35 artigos do repositório *Scopus*, 15 artigos do repositório *ScienceDirect* e 36 artigos do repositório *Google Scholar*.

O terceiro filtro gerou resultados diferentes quando executado de forma manual comparado ao resultados da IA generativa. Dos 121 artigos provenientes do segundo filtro, os autores selecionaram 50 que julgaram ser relevantes para a pesquisa, enquanto a IA generativa marcou 43 artigos. A matriz de

confusão apresentada na Figura 2 exibe o cruzamento entre os dados gerados pelos autores e as respostas fornecidas pela IA generativa.

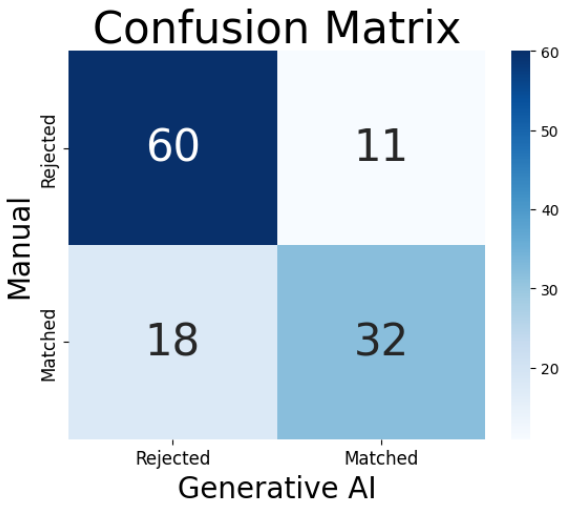


Figura 2. Matriz confusão do terceiro filtro.

Ao analisar a Figura 2, pode-se notar que a relação (r) da quantidade de artigos marcados como relevantes pelos autores e pela IA generativa foi de 74,42\%, valor obtido pela equação (1).

$$r = \frac{truePositive}{(truePositive + falsePositive)} \quad (1)$$

Já a concordância (c) entre as pesquisas rejeitadas pelos autores e pela IA é um pouco mais alta com 76,92\%, valor obtido pela equação (2).

$$c = \frac{trueNegative}{(trueNegative + falseNegative)} \quad (2)$$

No cenário geral quando se compara a concordância tanto dos arquivos marcados como relevantes quanto os rejeitados pelos autores e pela IA (cr), obtém-se o valor de 76,03\%, valor encontrado ao se aplicar a equação (3).

$$cr = \frac{(truePositive + trueNegative)}{total} \quad (3)$$

Com os resultados obtidos, avaliou-se a concordância entre os dois métodos de classificação utilizados, o manual e o gerado pela IA generativa. A questão central foi analisar os pontos de discordância entre os dois métodos buscando entender o porquê da divergência e tentando elencar qual dos métodos foi mais assertivo.

Ao analisar títulos que a inteligência artificial rejeitou, mas que os autores julgaram relevante para o tema, foi possível detectar alguns pontos onde a IA foi mais assertiva do que a classificação manual dos autores. Por exemplo, no título "*Low-light-level image super-resolution reconstruction via deep learning network*", por se tratar de uso de aprendizagem profunda para tratamento de imagens com baixo nível de luz, a priori pode parecer um trabalho muito relevante para o tema. Mas ao analisar com mais calma pode-se ver que os termos reconstrução de imagem de super resolução afastam este título do tema central do trabalho, por ser aplicado à uma tarefa muito específica (reconstrução) para um tipo de imagem também muito específico (imagem de super resolução).

Da mesma forma, a análise dos resultados revelou momentos em que os autores foram mais assertivos do que a IA. Por exemplo, o título "*The Deep Learning-Based Image Enhancement Method for High-Contrast Low-Light Images*", foi reprovado pela IA, provavelmente pelo fato de incluir o processamento de imagens com baixa iluminação e alto contraste. Porém os autores, com uma visão mais ampla sobre o estudo a ser desenvolvido, julgaram que este título poderia conter ferramentas muito úteis para sua revisão de literatura.

Está análise minuciosa sobre os pontos de discordância entre autores e IA, gerou novos resultados, aproximando a classificação de ambos os métodos como pode ser visto na Figura 3.

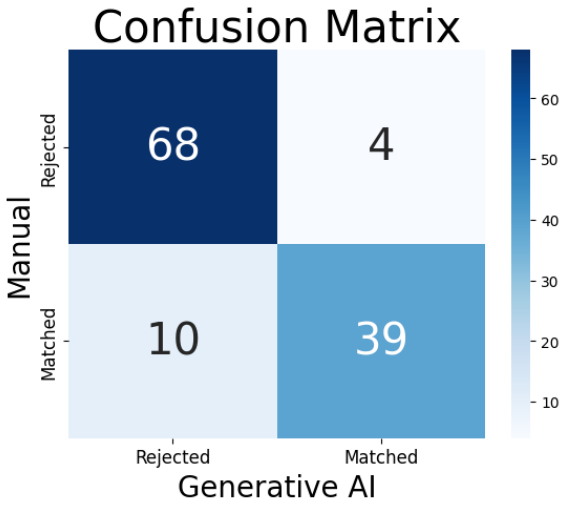


Figura 3. Matriz confusão do terceiro filtro após discussão dos resultados

## CONCLUSÃO

A metodologia empregada neste estudo trouxe excelentes resultados, além de se mostrar ser uma ferramenta extremamente útil ao pesquisador durante o desenvolvimento de sua revisão de literatura ou revisão bibliográfica.

Entretanto, a discussão dos resultados obtidos, evidenciou pontos de melhoria que podem ser explorados. Destaca-se nesse pontos, a análise das classificações geradas pela Inteligência Artificial onde constatou-se que, por falta de informações

adicionais sobre o tema tratado, a IA gerou uma classificação limitada.

Para contornar este problema, após a análise desenvolvida na discussão dos resultados, foi possível notar que se ao invés de se utilizar uma frase como padrão de comparação, fosse utilizado o resumo do trabalho a ser desenvolvido, a IA teria mais dados para analisar, o que provavelmente geraria uma classificação muito mais assertiva Indicar de forma objetiva as principais conclusões obtidas pelo trabalho.

## REFERÊNCIAS

Brizola, J. e Fantin, N. (2016). Revisão da literatura e revisão sistemática da literatura, Revista de Educação do Vale do Arinos-RELVA 3(2).

De Lima Andrade, E., da Cunha e Silva, D. C., de Lima, E. A., de Oliveira, R. A., Zannin, P. H. T. e Martins, A. C. G. (2021). Environmental noise in hospitals: a systematic review, Environmental Science and Pollution Research 28: 19629–19642.

De Sousa, L. M. M., Firmino, C. F., Marques Vieira, C. M. A., Severino, S. S. P. e Pestana, H. C. F. C. (2018). Revisões da literatura científica: tipos, métodos e aplicações em enfermagem, Revista portuguesa de enfermagem de reabilitação 1(1): 45–54.

Marconi, M. d. A. e Lakatos, E. (2005). Fundamentos de metodologia científica. (p. 165), São Paulo: Atlas.

Microsoft (2024). Conceitos básicos de ia generativa. Disponível em https://learn.microsoft.com/pt-br/training/modules/fundamentals-generative-ai/. Acesso em 30 de abril de 2024.

Miguel, M. L., dos Santos, L. J. and de Souza, L. A. M. (2022). Algumas percepções de estudantes do ensino médio sobre ciências, pseudociência e movimentos anticientíficos, Investigações em Ensino de Ciências 27(1): 191–222.

Rampazzo, L. (2002). Metodologia científica, Edições Loyola.