Miscitations in Scientific Papers: Dataset and Detection

Qinyue Liu[1,*], Amira Barhoumi[1] and Cyril Labbé[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

Abstract

Citations play an important role in scientific research. However, numerous inaccurate citations are found within scientific papers. These erroneous citations, also called miscitations, may result in the misinterpretation of research, distortion of the original author's intended meaning, and potentially even more serious consequences. The objective of our study is to detect automatically the erroneous citations, by assessing whether the citation contexts accurately reflects the content of the cited papers. We do not assess the correctness or formatting of the references within the reference list. We built a balanced "proof of concept" dataset, including both reliable and erroneous citations sourced from published open-access scientific papers, with the aim of investigating the feasibility of automatically detecting erroneous citations. In addition to the dataset, our study also introduces two methods for evaluating the reliability of citations using Natural Language Processing (NLP) techniques. The first method involves using cosine similarity metric. The second method implements a paraphrase classifier to distinguish between reliable and erroneous citations. Both methods take as input BERT embeddings and output the reliability of citations : reliable or erroneous. Based on our experimental results, the best performance is obtained with cosine similarity.

Keywords

Citation Reliability, Erroneous Citation, BERT, Textual Similarity, Misbehavior in Scientific Paper

1. Introduction

The use of citations in scientific papers is an essential practice, serving various purposes for authors. For instance, citations are utilized to establish research context, reference methodologies, or highlight contrasting findings or theories [1]. Additionally, citations also allow us to track the progression of idea and knowledge, as well as to assist readers in constructing a framework to build new hypothesis [2]. Nonetheless, various studies indicate that inaccurate citations are common in scientific papers. For example, one study mainly examining 28 papers reported a miscitation rate of 25.4% [3]. Another study focusing on journalistic papers, which reviewed a total of 49 articles, identified a citation error rate of 30.1% [4].

Inaccurate citations can lead to significant consequences. A study on the impacts of citations [2] discovered that over time, the most cited papers tend to become the conceptual symbol of a particular idea, and the lack of critical engagement leads certain studies to be cited much more frequently than their academic contribution warrants. According to the study, this phenomenon

Joint Proceedings of BIR 2024: 14th International Workshop on Bibliometric-enhanced Information Retrieval and IR4U2 2024: 1st Workshop on Information Retrieval for Understudied Users

*Corresponding author.

Qinyue.Liu@univ-grenoble-alpes.fr (Q. Liu); Amira.Barhoumi@univ-grenoble-alpes.fr (A. Barhoumi); Cyril.Labbe@imag.fr (C. Labbé)

not only harms the internal reward system of science but also diminishes epistemic diversity, impacting the breadth and variety of knowledge and perspectives within scientific discourse [2].

Some previous research have already begun to categorize miscitations. One particular study [5] has evaluated the contexts of citations referencing their preceding publication on focus group[6]. They defined two categories to describe citation errors:

- Citation contexts wherein authors employed descriptive information to substantiate their utilization of the focus group methodology.
- Citation contexts wherein authors consider the descriptive information on focus group in [6] as normative, or wherein the intention was not distinctly evident.

Another study [7] categorized citation errors into 3 groups: Trivial errors, slightly misleading errors, and serious errors. Inspired by these preceding studies, we defined two citation categories for our dataset in section 3.1.

Numerous prior studies have employed NLP techniques for citation analysis. The advent of large language models has significantly contributed to advancements in various citation analysis tasks. However, there's not much research focused on evaluating automatically the reliability of citations (Both recent articles evaluate manually the citations [8, 9]). In this regard, our study is directed towards automatically distinguish between citations that are reliable and those that are less so. The focus of our study is solely on assessing whether the context of citations accurately reflects the content of the cited papers. Our research does not extend to evaluating the correctness or formatting of the references within the reference list.

In this paper, we firstly review the related work. Secondly, we describe the collection of our datasets. Thirdly, we introduce our methods to distinguish reliable and erroneous citations using our own datasets. Fourthly, we present and discuss the results of evaluation. Finally, we conclude our work.

2. Related Work

In this section, we present various research related to citation analysis. Some of these research is centered on performing statistical analyses of erroneous citations within scientific papers. In contrast, other studies engage in analyzing citations using NLP techniques.

2.1. Statistical Analysis of Citations

In a study examining citation reliability through statistical methods, researchers classified citation errors into two types: major and minor. Major errors were defined as those that completely contradicted the claims of the cited authors, while minor errors were identified as inconsistencies or factual inaccuracies that did not fundamentally oppose the statements of the cited authors. Upon analyzing 27 articles, the study found that the median error rate for major errors was 11.5%, and for minor errors, it was 9.6%[3].
In another study of citation evaluation, the researchers evaluated 50 references, which were

randomly selected and published in OHNS[1] (Otolaryngology-Head and Neck Surgery)[10]. According to their analysis, citation errors occurred in 17% of studied references, of which 34% were classified as major. While these studies focus on evaluating citation reliability, some other statistical research [11, 12] intend to observe the themes in scientific papers through the citations. Some researchers led a study to analyze the context of citations in order to discover the trends and evolution of topics in biomedical and life sciences [11]. Another study introduced a methodology for analyzing and contrasting latent themes within scientific papers, specifically focusing on abstracts and sentences that reference a designated source paper [12].

2.2. Citation Analysis with NLP Techniques

Some other studies employing NLP techniques have engaged in various analytical tasks, including citation sentiment analysis [13] and citation polarity [14, 15] classification.

One study performed a sentiment analysis on the citations using a classifier based on the different embeddings[13]. In their work, the authors defined 3 sentiment classes for citations in his dataset, which is extracted from ACL (Anthology Reference Corpus [16]). They defined 3 sentiment categories of citation contexts: "N" implies negative sentiment, "P" implies positive, and "O" for objective. This sentiment analysis aims to distinguish citations according to these 3 classes.

Likewise, some other researcher studies aim to detect citation contexts criticizing cited sources[14, 15]. They fine-tuned various language models to classify critical and non-critical contexts. Using the CitaNeg and Critical Contexts Corpora[14], the researchers constructed their own corpus. Positive and neutral citations from CitaNeg were treated as non-critical, while those from Critical Contexts Corpus were seen as critical.

3. Experiments

We first collect both reliable and erroneous citations and used them to construct a dataset (Section 3.1), then we define two configurations (Section 3.2), for each configuration, we employ two classification methods for determining reliable and erroneous citations: cosine similarity method and paraphrase classifier method (Section 3.3).

A citation is considered unreliable if a justifying context is absent in the cited paper, or if the context in the cited paper does not support the citation. Conversely, a citation is considered reliable if it is supported by a context within the cited paper that justifies its use. In our experiments, we focus on assessing the similarity between the citation context in citing papers and the abstract section in the cited papers (Figure 1).

3.1. Data

This dataset is designed to evaluate the efficacy of our methodologies in categorizing citations as either reliable or erroneous. In our dataset, citations are "reliable" if they meet two criteria. First, they need to be in the same research domain as the cited paper. Second, they need to

[1]https://journals.sagepub.com/home/OTO

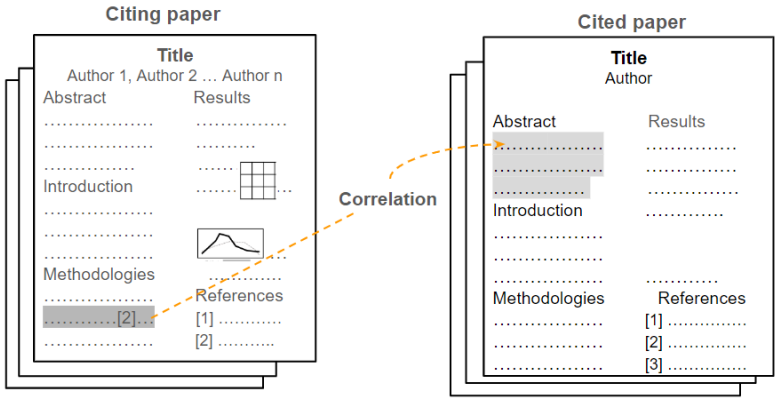


Figure 1: Detecting Miscitations Through Measuring Similarity between the Citation Context in the Citing Paper and the Abstract of the Cited Paper

accurately reflects the ideas of the corresponding cited paper. In contrast, "erroneous" citation lacks completely relevance to the ideas expressed in the cited paper, and it is also in a different research domain of the cited paper. Our data set is thus composed of very *favorable* real examples for which an automatic detection could have been used to prevent miscitations. Table 1 contains some examples of the categories we defined.

Table 1
Examples of the Erroneous and Out of Domain Citation and the Reliable and in Domain Citation in Our Dataset.

Category	Citation Context	Abstract in Cited Paper
Erroneous and out of domain	Eddy covariance devices or lysimeters can be used to determine ET0 (Vickers, 2017) [17].	Male moths compete to arrive first at a female releasing pheromone. A new study reveals that additional pheromone cues released only by younger females may prompt males to avoid them in favor of older but more fecund females [18].
Reliable and in domain	For instance, other approaches for topic modelling (e.g., Peinelt et al., 2020) can be tested [19].	Semantic similarity detection is a fundamental task in natural language understanding. Adding topic information has been useful for previous feature-engineered semantic similarity models, as well as neural models for other tasks [20]. ...

Citation contexts were manually gathered and annotated from various open-access papers that cited following 6 works: Most of the erroneous citations are extracted from the papers that referenced the work of (Payton et al.[21]), (Karthik et al.[22]), (Vickers et al.[18]). Conversely, the majority of reliable citations are extracted from the papers who cited the work of (Vaswani et al.[23]), (Peinelt et al.[20]), and (Glenton, Carlsen et al.[6]). To maintain dataset integrity, citing papers that were retracted were excluded, considering them as non-published works. In total, 199 citations were collected for the dataset. To ensure balance, 100 citations of them are

reliable and in domain, and 99 are erroneous and out of domain

Table 2
Number of Examples in Our Dataset for Miscitation Dectection.

Total Number	Reliable and in Domain	Erroneous and Out of Domain
199	100	99

3.2. Configurations

The abstract section of a scientific paper typically consists of multiple sentences. Considering that our citation context is composed of a single sentence, to perform different comparisons, we have created two configurations for the abstracts of the cited papers: the entire abstract configuration and the sliced abstract configuration. Throughout the experiment, for each configuration, we matched the citation contexts with the abstracts, defining these matched pairs as 'citation pairs'.

- **Entire Abstract Configuration**: In this configuration, we extract the entire text of the abstract section from the cited paper without making any modifications. We treat the whole abstract, which is composed of multiple sentences, as a single document. Then, we match this entire abstract with the citation context from the citing paper to measure their correlation.
- **Sliced Abstract Configuration**: Initially, the abstract section of the cited paper is segmented into individual sentences. Then, each sentence is paired with the citation context. For example, consider an abstract composed of 5 sentences. In this scenario, each sentence from the abstract is matched with the citation context from the citing paper. This process creates 5 distinct citation pairs. These pairs collectively represent the combined relationship between the citation context and the entire abstract. Using these pairs (citation context, sentence), we then evaluate the correlation between the citation context and the abstract of the cited paper.

3.3. Methods

For both configurations, we employed two methods to distinguish between reliable and erroneous citation: one based on the cosine similarity metric and the other based on a paraphrase classifier. Both methods take Bert embeddings of different documents as input (citation context, abstract or sentence).

- **Cosine Similarity**: it is a widely used method for measuring text similarity. Generally, the process begins by generating separate sentence embeddings for each text, capturing their semantic content. In our approach, we utilized the BERT model [24] to generate these sentence embeddings. Then, we apply the cosine similarity to assess the similarity between the two sentences.
  In the entire abstract configuration, we calculate directly the cosine similarity between the citation context and the entire abstract in the cited paper.

In the sliced abstract configuration, we calculate first the similarity between the citation context and each segmented sentence from the abstract, then we select the highest cosine similarity score from these comparisons to represent the overall relevance between the citation context and the abstract.
Cosine similarity values range from -1 to 1, where a value closer to 1 indicates bigger similarity between two texts, and a value closer to -1 indicates dissimilarity. Based on the *Receiver Operating Characteristic* (ROC)[25] curve analysis, we have established a threshold of 0.75 to differentiate between reliable and erroneous citations.

- **Paraphrase Classifier**: we fine-tuned our paraphrase classifier using BERT embeddings and the MSRP news corpus [26] to differentiate reliable and erroneous citations. The classifier's output is categorized as either 'paraphrase' or 'not paraphrase'. In our case, 'paraphrase' output means a reliable citation; otherwise, it is erroneous.
  In the entire abstract configuration, the classifier automatically categorizes each citation as either paraphrase (reliable) or not paraphrase (erroneous).
  In the sliced abstract configuration, we assess the proportion of the abstract that is classified as a paraphrase of the citation context. For example, if an abstract is segmented into 5 sentences and our classifier identifies 2 of these sentences as paraphrases of the citation context, we then infer that 40% (2 out of 5) of the abstract is considered a paraphrase of the citation context, and thereby reliable. Based on the analysis of the Receiver Operating Characteristic (ROC) [25] curve, we have established a threshold of 11%. This implies that if more than 11% of an abstract is identified as a paraphrase of the citation context, then the citation is considered reliable.

4. Results

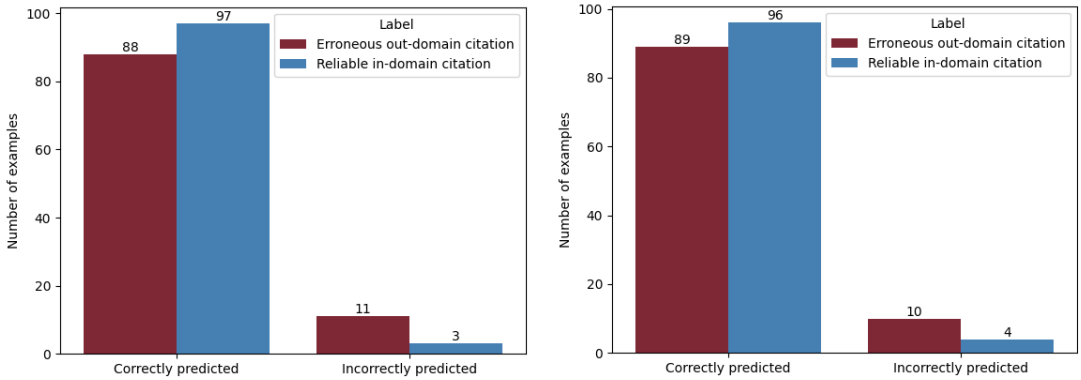
Table 3 presents the performance results of our methods. The cosine similarity method, in both the entire and sliced abstract configurations, has the best performance. Additionally, within the paraphrase classifier's applications, the sliced abstract configuration has a better performance compared to the entire abstract configuration.

Table 3
Experiment Results

Methods With Configurations	Accuracy	Accuracy Erroneous	Accuracy Reliable
Cosine Similarity Entire Abstract	93%	88.9%	97%
Paraphrase Classifier Entire Abstract	66.3%	87.9%	45%
Cosine Similarity Sliced Abstract	93%	89.9%	96%
Paraphrase Classifier Sliced Abstract	87.4%	89.9%	85%

Figure 2 illustrates that the cosine similarity method in both configurations achieved good performance, correctly predicting 185 out of 199 citations. According to ours results, this method demonstrates a stronger proficiency in accurately predicting reliable citations as opposed to erroneous ones.

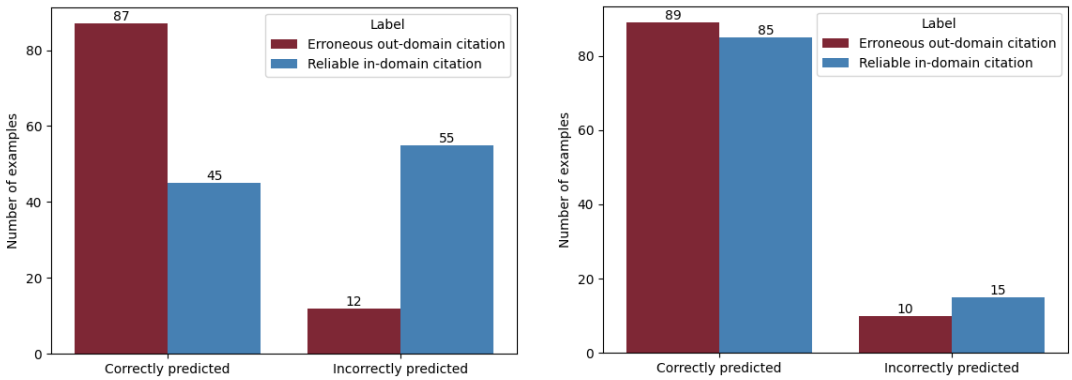
Figure 3 reveals that our paraphrase classifier correctly predicted 132 out of 199 citations with the entire abstract configuration. For the reliable citations, there are only 45 out of 100 correctly



(a) Entire Abstract Config (b) Sliced Abstract Config

Figure 2: Prediction Results for Cosine Similarity Method

predicted. However, it showed improvement in identifying erroneous citations, attaining a correction rate of just 87.9% in this aspect. When the classifier was applied to the sliced abstract configuration, we observed an improvement in predictive performance, with 174 out of 199 citations correctly predicted. Notably, this enhancement was accompanied by an increased ability to identify reliable citations.



(a) Entire Abstract Config (b) Sliced Abstract Config

Figure 3: Prediction Results for Classifier Method

Our results show that the cosine similarity method performs well with both configurations, achieving the best outcomes on our dataset. Regarding the classifier method, using the sliced abstract configuration enhances its effectiveness, especially in identifying reliable in-domain citations.

We dived deeper into the classification results. We analyze some wrongly classified cases

produced by both methods, and concluded the possible reasons for these misclassifications.

- The abstract of the cited paper is too short. In this scenario, with the abstract of the cited paper consisting of just two phrases, our paraphrase method in the sliced abstract configuration operates as follows: if at least one phrase is classified as a paraphrase of the citation context, this equates to 50% of the abstract being considered as a paraphrase. Consequently, the citation is immediately categorized as reliable. This is the main reason why some erroneous and out of domain citations are classified as reliable and in domain (Table 4).
- The supporting context in the cited paper justify the citation context in an indirect way. The classification error occurs when the content in the abstract indirectly supports the citation context. For instance, the citation context states *that deep learning has seen a rapid growth of increasingly capable and complex architectures*[27]. And the abstract of the cited paper indirectly demonstrates the capability of their architecture by showcasing its performance using the BLEU score. That is to say, the abstract does support the citation context, assuming that a high BLEU score implies a *capable architecture* (Table 5).
- Citing paper not totally out of domain with the cited paper. The cited paper reveals its objective to assess marine antibacterial extract for protease inhibitor activity and its anti-Plasmodium falciparum effects[22]. The citing paper[28], however, falls within the cancer research domain, and its citation context is unrelated to the cited paper's content. Despite this, the cosine similarity between the two is high. This could be attributed to the fact that both papers, though different in their specific research areas, belong to the broader biology domain. This shared domain may have contributed to the unexpected higher cosine similarity score, leading to a classification error (Table 6).

Table 4
An Incorrectly Classified Citation Example Caused by Too Short Abstract in the Cited Paper

Prediction	Label	Citation Context	Abstract in Cited Paper
Reliable	Erroneous	The first method is feature extraction/dimensional reduction, which transforms the original input feature into a reduced representation set. The second method is feature selection, which identifies relevant subsets while preserving the original information [10,11] [29].	Male moths compete to arrive first at a female releasing pheromone. A new study reveals that additional pheromone cues released only by younger females may prompt males to avoid them in favor of older but more fecund females[18].

Table 5
An Incorrectly Classified Citation Example Caused by Indirect Supporting Context in the Abstract

Prediction	Label	Citation Context	Abstract in Cited Paper
Erroneous	Reliable	Deep learning has witnessed an explosion of architectures of continuously growing capability and capacity [33, 25, 57][27].	... Our single model with 165 million parameters, achieves 27.5 BLEU on English-to-German translation, improving over the existing best ensemble result by over 1 BLEU. On English-to-French translation, we outperform the previoussingle state-of-the-art with model by 0.7 BLEU, achieving a BLEU score of 41.1 [23].

Table 6
An Incorrectly Classified Citation Example Caused by the Citing Paper Not Totally Out of Domain

Prediction	Label	Citation Context	Abstract in Cited Paper
Reliable	Erroneous	For example, the lentiviral method was used to target the CCR5 gene of T cells in humans for Cas9 delivery, but the main drawback of this method is the low knockout efficacies [186,187][28].	The study was planned to screen the marine actinobacterial extract for the protease inhibitor activity and its anti-Pf activity under in vitro and in vivo conditions. Out of 100 isolates, only 3 isolates exhibited moderate to high protease inhibitor activities on trypsin, chymotrypsin and proteinase K. ... The results confirmed up-regulation of TGF-b and down regulation of TNF-a in tissue and serum level in PbA infected peptide treated mice compared to PbA infection. The results obtained infer that the peptide possesses anti- Pf activity activity. It suggests that the extracts have novel metabolites and could be considered as a potential source for drug development[22].

5. Conclusion and Future Work

In this study, our main focus is to evaluate the reliability of citations in scientific papers, distinguishing between reliable and erroneous citations. To facilitate this, we constructed and annotated a dataset comprising 199 citation contexts, categorizing them into two groups: "reliable and in domain" and "erroneous and out of domain". Our research focuses on assessing the correlation between the citation context in the citing paper and the abstract section of the cited paper. We defined two configurations for this purpose: entire abstract and sliced abstract. Irrespective of the configuration, we utilized two methods: the first is based on cosine similarity, and the second employs a paraphrase classifier, with both methods using BERT embeddings as input. Our findings indicate that the sliced abstract configuration outperforms

the entire abstract configuration when employing the classifier method. The highest accuracy, reaching 93%, is achieved using the cosine similarity method in both configurations. However, our research has some limitations. We evaluated two citations in this paper using our cosine similarity method with the entire abstract configuration. The results are in Table 7.

Table 7
Evaluation of Two Citations from this Paper

Prediction	Label	Citation Context	Abstract in Cited Paper
Reliable	Reliable	Some researchers led a study to analyze the context of citations in order to discover the trends and evolution of topics in biomedical and life sciences [11].	With the exponential increase in the number of published papers, discovering how topics evolve ... This study proposes a large-scale analysis of the evolution of biomedical and life sciences using the citation contexts of the collected papers, or more precisely their citing sentences. Using 64,350 papers published in PubMed Central between 2008 and 2018, we determined the research trends for ten research topics. Moreover, we studied how these topics evolve across countries and across the most common journals in biomedical and life sciences [11].
Erroneous	Reliable	The citing paper[28], however, falls within the cancer research domain, and its citation context is unrelated to the cited paper's content.	The mechanisms involved in immune responses to cancer have been extensively studied for several decades, and considerable attention has been paid to harnessing the immune system's therapeutic potential. ... In this review, we have broadly focused on recent immunotherapeutic techniques against cancer and the use of CRISPR-Cas9 technology for the modification of the T cell, which can specifically recognize cancer cells and be used as immune-therapeutics against cancer [28].

For the first citation, our method correctly predicted it, because it's a typical reliable and in domain citation. However, our method failed to accurately predict the second citation. The reason for this failure is that the purpose of the second citation was to comment on the cited paper rather than to reflect its ideas. Consequently, while the citation is reliable, it is considered out-of-domain since our paper is in a different research field compared to the cited paper. This example underscores a limitation in our research. For now, our methods cannot assess citations similar to this example.

For future work, we plan to expand our dataset by introducing additional citation categories. Presently, our categorizations are limited to 'reliable and in domain' and 'erroneous and out of

domain' citations. However, there exist other types of citations not covered by these categories. An example of this, as demonstrated in Table 7, is a citation that does not fit into either of the existing categories in our dataset. The new citation categories are essential to rigorously test our existing methods and any potential new approaches. With an expanded dataset, we plan to divide it into training and testing sets to advance our research further. In addition to the model employed in our study, there exist more advanced language models that may be useful to investigate the task. We intend to explore these models to gain a deeper understanding of their potential impact. This exploration could provide valuable insights when applied to the analysis of citations.

Additionally, we are interested in conducting statistical research to determine the extent to which the abstract section of a cited paper supports the citation context in the citing paper. This analysis would help ascertain whether it's necessary to calculate correlations within the entire cited paper, or if focusing solely on the abstract section is sufficient.

Acknowledgments

We acknowledge the NanoBubbles project that has received Synergy grant funding from the European Research Council (ERC), within the European Union's Horizon 2020 program, grant agreement no. 951393.

Declaration on Generative AI

During the preparation of our work, we used ChatGPT in order to: Sentence Polishing. After using this tool, we reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Jurgens, S. Kumar, R. Hoover, D. McFarland, D. Jurafsky, Measuring the Evolution of a Scientific Field through Citation Frames, Transactions of the Association for Computational Linguistics 6 (2018) 391–406. doi:10.1162/tacl_a_00028.

[2] S. Horbach, K. Aagaard, J. W. Schneider, Meta-Research: How problematic citing practices distort science, MetaArXiv aqyhg, Center for Open Science, 2021. doi:10.31219/osf.io/aqyhg.

[3] B. C. Jergas H, Quotation accuracy in medical journal articles—a systematic review and meta-analysis, Peerj (2015).

[4] C. Kristof, Accuracy of Reference Citations in Five Entomology Journals, American Entomologist 43 (1997) 246–251. doi:10.1093/ae/43.4.246.

[5] C. Glenton, B. Carlsen, When "normal" becomes normative: A case study of researchers' quotation errors when referring to a focus group sample size study, International Journal of Qualitative Methods 18 (2019) 1609406919841251. doi:10.1177/1609406919841251.

[6] C. Glenton, B. Carlsen, What about n? a methodological study of sample-size reporting in focus group studies, BMC Med Res Methodol (2011). doi:10.1186/1471-2288-11-26.

[7] G. De Lacey, C. Record, J. Wade, How accurate are quotations and references in medical journals?, Br Med J (Clin Res Ed) 291 (1985) 884–886.
[8] V. Pavlovic, T. Weissgerber, D. Stanisavljevic, T. Pekmezovic, O. Milicevic, J. M. Lazovic, A. Cirkovic, M. Savic, N. Rajovic, P. Piperac, et al., How accurate are citations of frequently cited papers in biomedical literature?, Clinical Science 135 (2021) 671–681.
[9] A. Agarwal, M. Arafa, T. Avidor-Reiss, T. A.-A. A.-M. Hamoda, R. Shah, Citation errors in scientific research and publications: Causes, consequences, and remedies, The World Journal of Men's Health 41 (2023).
[10] M. F. e. a. Armstrong, Reference errors in otolaryngology-head and neck surgery literature, Otolaryngology–head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery (2018). doi:10.1177/0194599818772521.
[11] C. M. J. Jebari C, Herrera-Viedma E, The use of citation context to detect the evolution of research topics: a large-scale analysis, Scientometrics (2021).
[12] S. Liu, C. Chen, The differences between latent topics in abstracts and citation contexts of citing papers, Journal of the American Society for Information Science and Technology 64 (2013) 627–639. doi:https://doi.org/10.1002/asi.22771.
[13] H. Liu, Sentiment analysis of citations using word2vec, CoRR abs/1704.00177 (2017). arXiv:1704.00177.
[14] S. Te, A. Barhoumi, M. Lentschat, F. Bordignon, C. Labbé, F. Portet, Citation Context Classification: Critical vs Non-critical, Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022.
[15] F. Bordignon, Critical citations in knowledge construction and citation analysis: from paradox to definition, Scientometrics 127 (2022) 959–972.
[16] S. Bird, R. Dale, B. Dorr, B. Gibson, M. Joseph, M.-Y. Kan, D. Lee, B. Powley, D. Radev, Y. Tan, The acl anthology reference corpus: A reference dataset for bibliographic research in computational linguistics., 2008.
[17] H. Talebi, S. Samadianfard, K. Valizadeh Kamran, Investigating the roles of different extracted parameters from satellite images in improving the accuracy of daily reference evapotranspiration estimation, Applied Water Science 13 (2023). doi:10.1007/s13201-022-01862-6.
[18] N. J. Vickers, Animal communication: When i'm calling you, will you answer too?, Current Biology 27 (2017) R713–R715. doi:https://doi.org/10.1016/j.cub.2017.05.064.
[19] F. Chiarello, L. Gastaldi, A. Martini, Design and implementation of a text mining-based tool to support scoping reviews, International Journal of Technology Management 91 (2023) 147–161. doi:10.1504/IJTM.2023.128799.
[20] N. Peinelt, D. Nguyen, M. Liakata, tBERT: Topic Models and BERT Joining Forces for Semantic Similarity Detection, Association for Computational Linguistics, Online, 2020. doi:10.18653/v1/2020.acl-main.630.
[21] E. Payton, J. Khubchandani, A. Thompson, J. H. Price, Parents' expectations of high schools in firearm violence prevention, Journal of community health 42 (2017) 1118–1126.
[22] L. Karthik, G. Kumar, T. Keswani, A. Bhattacharyya, S. S. Chandar, K. Bhaskara Rao, Protease inhibitors from marine actinobacteria as a potential source for antimalarial compound, PloS one 9 (2014) e90972.
[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser,

I. Polosukhin, Attention is all you need 30 (2017).
[24] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). arXiv:1810.04805.
[25] J. A. Hanley, et al., Receiver operating characteristic (roc) methodology: the state of the art, Crit Rev Diagn Imaging 29 (1989) 307–335.
[26] W. B. Dolan, C. Brockett, Automatically constructing a corpus of sentential paraphrases (2005).
[27] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, R. Girshick, Masked autoencoders are scalable vision learners (2022) 16000–16009.
[28] M. Al Saber, P. Biswas, D. Dey, M. A. Kaium, M. A. Islam, M. I. A. Tripty, M. H. Rahman, T. I. Rahaman, M. Y. Biswas, P. Paul, M. A. Rahman, M. N. Hasan, B. Kim, A comprehensive review of recent advancements in cancer immunotherapy and generation of car t cell by crispr-cas9, Processes 10 (2022). doi:10.3390/pr10010016.
[29] Y. Gebreyesus, D. Dalton, S. Nixon, D. De Chiara, M. Chinnici, Machine learning for data center optimizations: Feature selection using shapley additive explanation (shap), Future Internet 15 (2023). doi:10.3390/fi15030088.