International Journal of Information Studies and Libraries
9 (1) 2024, 63-79
http://publishingindia.com/ijisl/

Status of Contrasting Citations in Funded Research

Sarita Gulati*, Babita Yadav, Atasi Sinhababu***, Rupak Chakravarty******

Abstract

This study attempts to analyse the context of the citations with the help of an AI and deep learning model-based web tool, Scite. The study analyses contrasting, supporting, mentioning, Lifetime SI and total cites metrics of citation statements and also examines the relationship between Lifetime SI and total cites, total cites and contrasting; total cites and supporting and contrasting and supporting. The data for this research has been derived from Scite's Funders citation data and was sorted based on contrasting citation records by the Funding Institutions in descending order for meaningful inferences. A correlation test was applied using the open-source statistical software Jamovi. Based on an analysis of citation statements related to research productivity, it was found that the "National Natural Science Foundation of China" has the highest number of publications in each category of editorial notices. Lifetime SI and total cites exhibit a weak positive correlation. In contrast, the parameters total cites and contrasting, total cites and supporting and contrasting and supporting show a significant and strong positive correlation.

Keywords: Scite, Citation Statements, Artificial Intelligence, Deep Learning, Citations, Machine Learning, Funders/Funding Institutions

Introduction

Contrasting research is a method of analysis that involves bringing together multiple studies on a particular topic with similar but different findings. The goal of contrasting research is to identify the studies' patterns, similarities and differences to gain a deeper understanding of the topic. One of the critical benefits of contrasting research is that it allows researchers to identify inconsistencies or contradictions in previous research. By analysing these inconsistencies, researchers can develop new hypotheses or theories that can help explain the discrepancies. It also helps to identify areas where further research is needed. For example, if multiple studies have conflicting findings on a particular topic, this may indicate that more research needs to be done to clarify the issue (Chiong et al., 2019).

Science is a systematic and logical approach to discovering how the natural world works. Science is a broad field that encompasses many different disciplines, including biology, chemistry, physics, geology, astronomy and more. Each of these disciplines has unique methods and approaches to scientific inquiry, but they all share a commitment to rigorous testing and empirical evidence (The Process of Science, 2021). So, Science can be defined as the process of thinking, using observation,

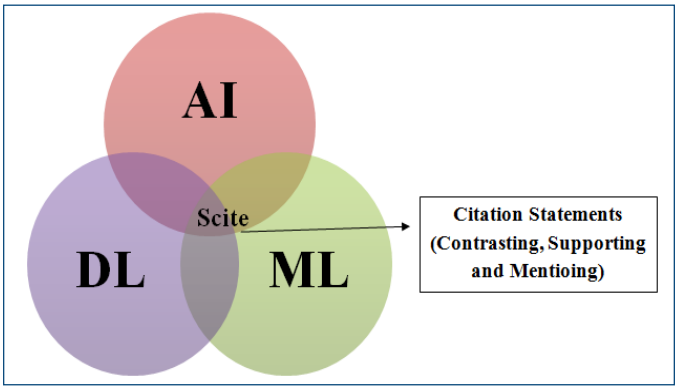
* Librarian, Shivalik Institute of Education & Research, Mohali, Punjab, India. Email: gulatisarita2005@gmail.com; ORCiD: https://orcid.org/0000-0001-9963-8917
** Library Assistant, Central university of Punjab, Bathinda, Punjab, India. Email: babitayadav112@gmail.com; ORCiD: https://orcid.org/0000-0002-1905-0863
*** Assistant Librarian, Amity University, Mohali, Punjab, India. Email: asinhababu@pb.amity.edu; ORCiD: https://orcid.org/. 0000-0001-7036-0827
**** Department of Library & Information Science, Panjab University, Chandigarh, India. Email: rupak@pu.ac.in; ORCiD: https://orcid.org/0000-0001-5046-1663

inferring, comparing, contrasting and looking for causes and effects. It involves formulating hypotheses based on observations and testing them through experiments or further observations. The results of these tests are then used to refine or reject the original hypothesis, leading to new knowledge and understanding.

The growth and progress we witness today leading to the transition from a nomadic society to an information society can be attributed to scientific advancements. The development of science has played a critical role in shaping human society. Advancements embedded in the scientific literature, including patents and research publications, could not have been achieved in isolation as novelty, innovation and inventions have always been inspired by previous research and discoveries. Citations are essential to academic writing that acknowledges previous scholarly works and demonstrates credibility. Proper citation practices help maintain academic integrity and avoid plagiarism. Citations provide a way for authors to give credit to the work of others and demonstrate the credibility of their research. There are different citation styles, including APA, MLA, Chicago and Harvard, among others, and each style has specific guidelines for formatting citations in academic papers. The Web of Science (WoS) and Scopus both serve the purpose of indexing and abstracting (A&I) databases and citation indexing. Citation indexes are used for citation analysis. Citation analysis is a method used to evaluate the impact and importance of scholarly research by analysing the number of times an article, book or other publication has been cited by other researchers in their work. Citation analysis has become increasingly important in academia, as it provides a way to evaluate the quality and impact of research objectively. Funding agencies and academic institutions often use it to make decisions about grants, promotions and tenure. Citations are very relevant indicators of disagreements and agreements in creating scientific knowledge. New knowledge/ innovation depends on agreeing on ideas/concepts having a universal consensus, contrasting ideas and conflicts and disagreements in the process of knowledge creation.

Citation indexes previously only displayed links between scientific papers without providing further context, such as why a citation was made. Because of the lack of context and limited metadata available beyond paper titles, authors and the date of publication, it has only been possible to calculate how many times a work has been cited, not to analyse its citations, whether they are intended to support another publication, refute it, provide methodological support or serve as an example. This needs to be revised, given citations' central role in research evaluation. Undoubtedly, not every citation is made equally, yet we have treated them all the same. It is for this reason that scientific controversy exists. There are always controversies in science. Contrasting opinions on specific topics can sometimes be called a "persistent disagreement" over knowledge. A scientific controversy is a substantial disagreement among scientists. It involves issues such as the interpretation of data, which ideas are most supported by evidence, and which ideas are the most worth pursuing (Egger & Carpi, 2010).

The Research Literature is an essential resource for summarising the critical findings of a study. However, research claims on the same topic can be contradicting. Contradictory claims answer the question with conflicting assertion values (Yes or No). An innovative web tool, scite (scite.ai), was used to identify contradictory research claims and analyse the context of the citations in this study. Scite a new citation index and tool that uses recent advances in artificial intelligence (AI) to produce "Smart Citations." Smart Citations reveal how a scientific paper has been cited by providing the citation context and a classification system describing whether it provides supporting or contrasting evidence for the cited claim or just mentions it (Nicholson et al., 2021).



Scite: AI (Artificial Intelligence), DLM (Deep Learning Model) and Machine learning (ML)

Scite (https://scite.ai) is a platform that offers a unique approach to scientific research by analysing and verifying the credibility of scientific articles. It allows users to

search for and see how publications have been cited. Other citation indexes let us know what publications cite one another but not how they cite it, or read all those mentions of another publication in one place. For this, the Scite extract classifies and presents the context of the citation made to a particular article, known as citation statements. Citation Statement (citance) is the surrounding context of a citation when the citation is made. This includes the sentences surrounding the citation, which section the citation was made in, whether the citation is meant to support or contrast an article, and more (Rosati, 2021). The Scite classifies citations based on the rhetorical function of a citation: whether the citation supports, disputes or mentions another article. The algorithm of this classification is based on AI, Deep Learning Model (DLM) and Machine Learning (ML). It has classified 8.23 billion citation statements.

While Scite does not provide funding directly, it offers valuable insights into funding institutions that support scientific research. There are various types/sources of funding institutions in Scite, including government agencies, private foundations and corporations. These institutions offer grants, scholarships and fellowships and play a significant role in providing financial support for researchers across different fields.

AI, ML and DLM

Artificial Intelligence (AI) is the field of research where mathematicians, computer scientists and software engineers design, develop and test techniques for making AI systems more transparent and comprehensible by their stakeholders. Most approaches in this field require technical expertise to manipulate the recursive algorithms that implement the mathematical functions at the roots of deep learning (Futia & Vetrò, 2020). According to TechTarget, AI is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition and machine vision. AI leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind. The Encyclopaedia Britannica states, "Artificial intelligence (AI) is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings." Using recent advances in AI, Scite creates "Smart Citations".

Machine learning (ML) is a branch of AI and computer science that focuses on using data and algorithms to imitate how humans learn, gradually improving its accuracy (What Is Machine Learning? | IBM, 2016). ML is a subset of AI that enables computer systems to learn and improve from experience without being explicitly programmed. It is an application of statistical models and algorithms that allow machines to automatically learn patterns and relationships in data and use them to make predictions or decisions.

Deep Learning (DL) attempts to mimic the human brain—albeit far from matching its ability—enabling systems to cluster data and make predictions with incredible accuracy. DL drives several AI applications and services that improve automation, performing analytical and physical tasks without human intervention (ibm.com, 2015). DL is a subfield of ML concerned with algorithms inspired by the structure and function of the brain, called artificial neural networks (Brownlee, 2019). DL is an essential element of data science, which includes statistics and predictive modeling. It is highly beneficial to data scientists tasked with collecting, analysing and interpreting large amounts of data; DL makes this process faster and easier (Burns & Brush, 2021). Scite provides a way to categorise the different citation types as per their use & requirement in the study. The DLM of scite classifies a citation as supporting, contradicting or simply mentioning to add quality value to the citation counts (Hayes, 2020). This citation quality is measured by the Scite journal index (sji), calculated as,

$$SI = \frac{\#\ Supporting\ Cites}{\#\ Supporting\ Cites + \#\ Contrasting\ Cites}$$

Study Background

Scite (scite.ai) ingests scientific papers from various sources, including publishers, open-access articles and other sources. It is a new citation index and tool that takes advantage of recent advances in AI to produce "Smart Citations. Recently, Nicholson et al. (2021) made a significant contribution to the annotation and classification of "supporting" v.s. "contrasting" relationships. Nicholson et al. (2021) applied ML, traditional document ingestion methods and a network of researchers to develop a "smart citation index" called Scite, to categorise citations

based on context. Through various examples, the author shows how a citation is used, presenting the surrounding textual context from the citing paper and applying a DLM to determine if a statement provides supporting or contrasting evidence for the referenced work or mentions it. Through various illustrations, it describes Scite work and how it can be used to further research and evaluation. It uses real-world examples of citation classifications to explain why a citation type has or has not been assigned. Rife et al. (2021) in their study recognised the importance of citation contexts and discussed Scite features like the Scite report, which displays all citations of a given article, and can be filtered by publication year and type, as well as the type of citation-supporting, mentioning or disputing—as classified by Scite's ML algorithm. The authors also discuss various implications of smart citations. Scite also displays metadata about a given article retrieved daily from Crossref, PubMed and DataCite. One key piece of information extracted is editorial notices. Uppala et al. (2022) discusses a set of methods for detecting editorial notices and a novel approach that uses metadata mining to detect notices that are not explicitly classified as notices. Authors produce a centralised and open compilation of Digital Object Identifiers (DOIs) to which editorial notices have been attached. Several systems have been developed to track these notices in the last few years, from RetractionWatch (https://retractionwatch.com/), which manually curates a list of retracted articles, to Crossref and Pubmed, which take updates from publishers and other sources, and provide that data on their platforms. However, how this information is communicated has always been inconsistent, with different publishers taking various approaches to broadcast their notices. Few publishers rename the article "Retracted"; others add prefixes or suffixes to the original titles. In contrast, others issue a new notice DOI indicating an existing article is retracted without any indication of the original one scite (Scite.ai, 2022).

The Scite corpus classified types of citations, including supporting, mentioning and contrasting. It also displays the sections of citations, such as whether the citation occurs in the introduction, methods, results or discussion sections. Scite's classifications of citations as supporting or contradicting can help researchers understand if a particular paper's traditional citation indices have been manipulated or are biased since this index is much less susceptible to manipulation or bias. It can also provide more excellent value to publishing non-significant results since non-significant results could later be contradicted or supported accordingly, and help to remove the strong positive result so prevalent in scientific research.

Study Objectives

Scite classifies the citation categories as contrasting, supporting and mentioning utilising AI, DLM and ML functionality. The present study attempts to analyse the status of citation statements about the research productivity of select funding institutions using Scite. The specific objectives of the study are:

OB1: To analyse total cites and the total number of articles published by funding institutions.

OB2: To examine contrasting, supporting and mentioning metrics associated with citation statements.

OB3: To investigate the status of the Lifetime SI (Scite Index) of citation statements.

OB4: To explore the correlation between Lifetime SI with total cites, contrasting citations and supporting citations, respectively.

OB5: To explore the correlation between contrasting and supporting.

OB6: To assess the status of retracted, withdrawn, corrected and "expression of concern" about research publications of the funding institutions.

Hypotheses

Generalisation is one of the most crucial aspects of the scientific process leading to the creation of new knowledge. The data extracted from the scite database were tested empirically through hypotheses testing involving correlation analysis between study variables associated with the citation statements. The statistical tests were performed using the open-source software Jamovi for inferential statistics.

Ho1:	No significant relationship exists between Lifetime SI and total cites received by funding institutions.
Ho2:	No significant relationship exists between total citations and contrasting citations.
Ho3:	There is no significant correlation between total citations and supporting citations.
Ho4:	No significant association exists between contrasting and supporting citations.

Research Methodology

In order to achieve the study objectives, the data available in the scite database were analysed. The responses recorded by the scite platform were categorised primarily into two classes: supporting and disputing evidence and information about "mentions." Scite shows how a citation was used by displaying the surrounding textual context from the citing paper and classification from our DLM that indicates whether the statement provides supporting or contrasting evidence for a referenced work or mentions it. Scite has ingested papers from over 32 million articles, and the scite database contains over 1.2 billion citation statements (Enago Academy, 2021). The primary data for this research has been derived from the Scite's funding institutions' citation data, which includes funders' names, Lifetime Scite Index (SI), total citations, supporting citations, mentioning citations and contradicting citations. The whole data was sorted based on contrasting citation records by the organisation in descending order for meaningful inferences.

The relevant data for statistical analysis involving Correlation Analysis (CA) and testing of hypotheses were extracted during the third week of March 2023 from Scite (https://scite.ai). A total of 5000 records were exported from the portal to an Excel file for further analysis and investigation. To make the findings of this paper reproducible, an Excel sheet/tables of the funding institutions along with all of their citation data are prepared. Data cleansing was performed, and data were processed with the help of the open-source statistical package Jamovi (https://www.jamovi.org/) for inferential statistics, including correlation tests. Pearson's correlation test was applied and decisions were taken based on Pearson's correlation coefficient (r) and the p-value results. The correlation test was applied to test the nature of the association between Lifetime SI and total cites, total cites and contrasting, total cites and supporting and contrasting and supporting. For plotting the word cloud, the service at https://www.wordclouds.com was used.

Results

Table 1: Contrasting Citation Statement of Top 20 Funding Institutions

Sr. No.	*Funders Name*	*Total Cites (Millions)*	*Contrasting (Thousands)*	*Supporting (Thousands)*	*Mentioning (Millions)*	*Lifetime SI*	*Total Publications (Thousands/ Millions)*
1	National Institute of Health (NIH, USA)	15.2M	72.8K	689K	14.2M	0.9	480K
2	United States Department of Health and Human Services (HHS, USA)	13.9K	68K	637K	13.1M	0.9	370K
3	National Natural Science Foundation of China (NSFC, China)	20.8K	61K	678K	19.8M	0.92	2.0M
4	National Science Foundation (NSF, USA)	7.7K	27.1K	313K	7.3M	0.92	320K
5	European Commission (Belgium)	7.1K	26.2K	306K	6.7M	0.92	280K
6	Deutsche Forschungsgemeinschaft (DFG, Germany)	4.1K	16.5K	195K	3.8M	0.92	190K

Sr. No.	Funders Name	Total Cites (Millions)	Contrasting (Thousands)	Supporting (Thousands)	Mentioning (Millions)	Lifetime SI	Total Publications (Thousands/ Millions)
7	United States Department of Energy, USA	4.5M	14.9K	201K	4.2M	0.93	140K
8	Ministry of Education, Culture, Sports, Science and Technology, Japan	2.7M	12.7K	130K	2.5M	0.91	120K
9	European Research Council (ERC, Belgium)	3.3M	10.9K	142K	3.1M	0.93	110K
10	Ministry of Science and Technology of the People's Republic of China (China)	3.3M	9.6K	112K	3.2M	0.92	190K
11	Natural Sciences and Engineering Research Council (NSERC, Canada)	2.2M	9.1K	92.8K	2.1M	0.91	130K
12	United States Department of Defense (USA)	3.1M	9.1K	107K	3.1M	0.92	94K
13	Japan Society for the Promotion of Science (JSPS, Japan)	2.1M	8.9K	95.7K	1.9M	0.91	160K
14	National Institute of Mental Health (NIMH, USA)	1.1M	8.5K	69.3K	1.1M	0.89	36K
15	Ministry of Economy, Industry and Competitiveness (Spain)	1.7M	6.9K	82.5K	1.6M	0.92	90K
16	Swiss National Science Foundation (SNF, Switzerland)	1.6M	6.7K	75.1K	1.5M	0.92	57K
17	Australian Research Council (ARC, Australia)	1.7M	6.6K	67.7K	1.6M	0.91	76K
18	National Cancer Institute (USA)	1.5M	6.3K	62.9K	1.5M	0.91	55K
19	National Research Foundation of Korea (NRF, South Korea)	2M	6.2K	67.8K	1.9M	0.92	170K
20	National Council for Scientific and Technological Development (Brazil)	1.3M	6.2K	62.2K	1.1M	0.91	120K

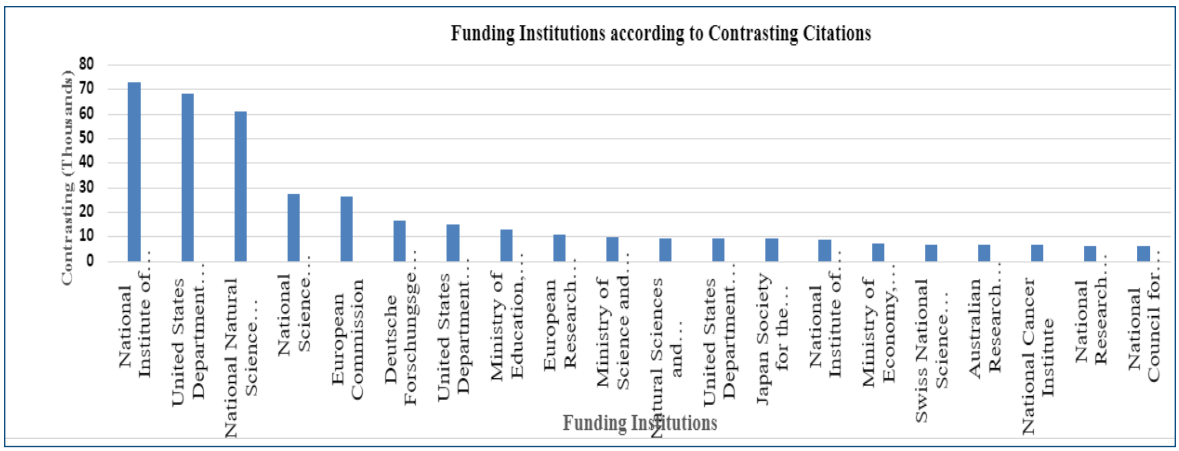


Fig. 1: Top 20 Funding Institutions According to Contrasting Citations

Table 1 of funding institutions' Smart Citation metrics depicts information on contrasting citation statements of the Top 20 funding institutions along with total cites, Lifetime SI and publications. Total citations refer to the total number of times that the research funded by the funder has been cited in other scientific publications. The higher the number of citations, the more impactful the research is considered to be. Contrasting and supporting mentions refer to the number of times that the research funded by the funder was cited in a contrasting or supporting manner. Contrasting mentions are those that dispute or contradict the research, while supporting mentions agree with and support the research. Lifetime SI refers to the number of publications that the funder has supported throughout its lifetime.

Fig. 2: Visualization of Research Areas/Disciplines Covered in Funded Research/Disciplinary Coverage of Contrasting Research

This table help to discover how many contrasting citations of mentioned funding institutions have received. This table mentions that the National Institutes of Health has the highest contrasting citation, i.e., 72.8K, with a total citation of 480K, followed by "The United states department of Health and human services," which has a contrasting citation of 68K. "National Natural Science Foundation of China" has contrasting citations of 61K with the highest Publication of 2.0M. "National Research Foundation of Korea" and " National Council for Scientific and Technological Development" have the lowest contrasting citation, i.e., 6.2K.

Each discipline is represented with different colours and sizes depending on the disciplinary areas funded institutions cover (Fig. 2). Thus, it can be seen that science and technology disciplines received maximum funds from funding institutions, and their contrasting research visibility is more followed by education, biomedical and health. The domains that receive fewer funds are, for example, agriculture, aerospace, arts and humanities, followed by physics, law, innovation, environment and food.

Table 2: Top 5 Funding Institutions According to Lifetime SI

Sr. No.	*Funders Name*	*Lifetime SI*	*Total Cites*	*Contrasting*	*Supporting*	*Mentioning*	*Total Publications*
1	COMSATS University Islamabad	0.99	2,159	1	108	2,017	192
2	European Investment Bank	0.99	1,983	1	101	1,852	69
3	National Institute for Fusion Science	0.98	2,542	3	181	2,351	631
4	Comisión Nacional De Energía Atómica	0.98	3,449	3	169	3,246	248
5	Marshall Space Flight Center	0.98	4,566	4	218	4,329	219

Table 2 gives information about the Top five funding institutions according to Lifetime SI. The COMSATS University Islamabad, with a lifetime SI score of 0.99, is listed at the top of the table. This implies that research papers published by this university have received higher citation rates compared to the average number of citations received by similar papers in the same field. The university has 2159 citations, of which 108 are contrasting, 2017 are supporting and 1, in this case, is mentioned. "COMSATS University Islamabad" and "European Investment Bank" both have the highest Lifetime SI (0.99), and the rest of the last three have the same Lifetime SI (0.98).

Table 3: Top 5 Funding Institutions According to Total Cites

Sr. No.	*Funders Name*	*Total Cites (Millions)*	*Contrasting (Thousands)*	*Supporting (Thousands)*	*Mentioning (Millions)*	*Lifetime SI*	*Total Publications (Millions/ Thousands)*
1	National Natural Science Foundation of China	20.8M	61.2K	678K	19.8M	0.92	2.0M
2	National Institutes of Health	15.2M	72.8K	689K	14.2M	0.9	480K

Sr. No.	*Funders Name*	*Total Cites (Millions)*	*Contrasting (Thousands)*	*Supporting (Thousands)*	*Mentioning (Millions)*	*Lifetime SI*	*Total Publications (Millions/ Thousands)*
3	United States Department of Health and Human Services	13.9M	68.1K	637K	13.1M	0.9	370K
4	National Science Foundation	7.7M	27.1K	313K	7.3M	0.92	320K
5	European Commission	7.1M	26.2K	306K	6.7M	0.92	280K

The table shows that the "National Natural Science Foundation of China" is the top funder in terms of total citation count, with over 2.0 million citations. However, the National Institutes of Health, the United States Department of Health and Human Services, and the National Science Foundation are also among the top funders with high citation counts and significant supporting and mentioning publications. The European Commission is the fifth-ranked funder in terms of total citation count, with over seven million citations, but has a lower number of supporting publications and mentions compared to the other funders in the top five.

Based on total cites, "The National Natural Science Foundation of China" has secured the maximum number of citations, 20.8M cites (publications =2.0M), followed by the National Institutes of Health with 15.2M (publications =480 K) and the United States Department of Health and Human Services with 13.9M (publications =370K).

Table 4: Editorial Notices According to Contrasting Citations (Top 10 Funding Institutions)

Sr. No.	*Funders Name*	*Retraction*	*Withdrawals*	*Corrections*	*Errata*	*Expression of Concern*	*Total Publications (Thousands/ Millions)*
1	National Institutes of Health	228	58	3246	6696	8	480K
2	United States Department of Health and Human Services	210	47	2797	5809	7	370K
3	National Natural Science Foundation of China	1386	314	6058	9563	49	2.0M
4	National Science Foundation	54	16	1541	2254	2	320K
5	European Commission	28	21	1566	2313	3	280K
6	Deutsche Forschungsgemeinschaft	29	10	1200	1545	2	190K
7	United States Department of Energy	19	15	696	1127	0	140K
8	Ministry of Education, Culture, Sports, Science and Technology	52	10	247	533	2	120K
9	European Research Council	9	7	690	989	1	110K
10	Ministry of Science and Technology of the People's Republic of China	69	20	413	1099	3	190K

Scite provides a count of editorial notices for each article. Journal editors publish editorial notices to notify readers about retractions, corrections, errata, withdrawals and expressions of concern. Retraction is a mechanism for correcting the literature and notifying readers of significant concerns about an article's integrity, validity or reliability. Withdrawals are the same as retraction, but (often, but not always) initiated by the author(s) themselves. Published works will be withdrawn if there is evidence of unethical research, unreliable data, misconduct or plagiarism. *Corrections* are notices about an aspect of a published work that is otherwise valid and should therefore remain in the scientific record (Wager et al., 2009). They address errors, but there are no concerns about the integrity or reliability of the reported work. Errata corrects a significant error in a published text (Elsevier, 2016). Expressions of Concern (EOCs) are notices published at editors' discretion to alert readers to severe concerns about published work or an article's compliance (Plos.org, 2023).

Table 4 depicts the list of the top 10 funding institutions based on contrasting. Each funding institution shows the number of retractions, corrections, errata, withdrawals and concerns. The National Institutes of Health and the United States Department of Health and Human Services, both based in the US, are ranked 1st and 2nd, respectively. These organisations have a significant number of retractions, withdrawals, corrections, errata and expressions of concern, and have published 480K and 370K articles respectively. The number of papers in the Errata category is high (9563), while the number of articles in the Expressions of Concern category is lower (8). The National Natural Science Foundation of China, with a total of 2 million publications, has the highest number of retractions (1386) and corrections (6058) among the listed organisations, and also has a high number of withdrawals (314) and expressions of concern (49). Also, it indicates that retractions, corrections, errata, and withdrawals increase with higher publication rates. Overall, the table provides insight into the quality of research and publication practices of various funding Institutions.

Correlation Analysis (CA)

Correlation analysis is a statistical technique used to measure the strength of the relationship between two or more variables. In essence, correlation analysis helps to determine whether and to what extent two variables are related. The degree of association is measured by a correlation coefficient, denoted by r. The strength of the relationship is measured by a correlation coefficient, which ranges from -1 to +1. A correlation coefficient of +1 indicates a perfect positive relationship, while a correlation coefficient of -1 indicates a perfect negative relationship.

Table 5: Correlation Analysis between Lifetime SI and Total Cites

Correlation Matrix			
		Lifetime SI (LSI)	*Total Cites (TC)*
Lifetime SI	Pearson's r	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	N	—	
Total Cites	Pearson's r	0.030	—
	p-value	0.073	—
	95% CI Upper	0.063	—
	95% CI Lower	-0.003	—
	N	3488	—

Note. * p < .05, ** p < .01, *** p < .001

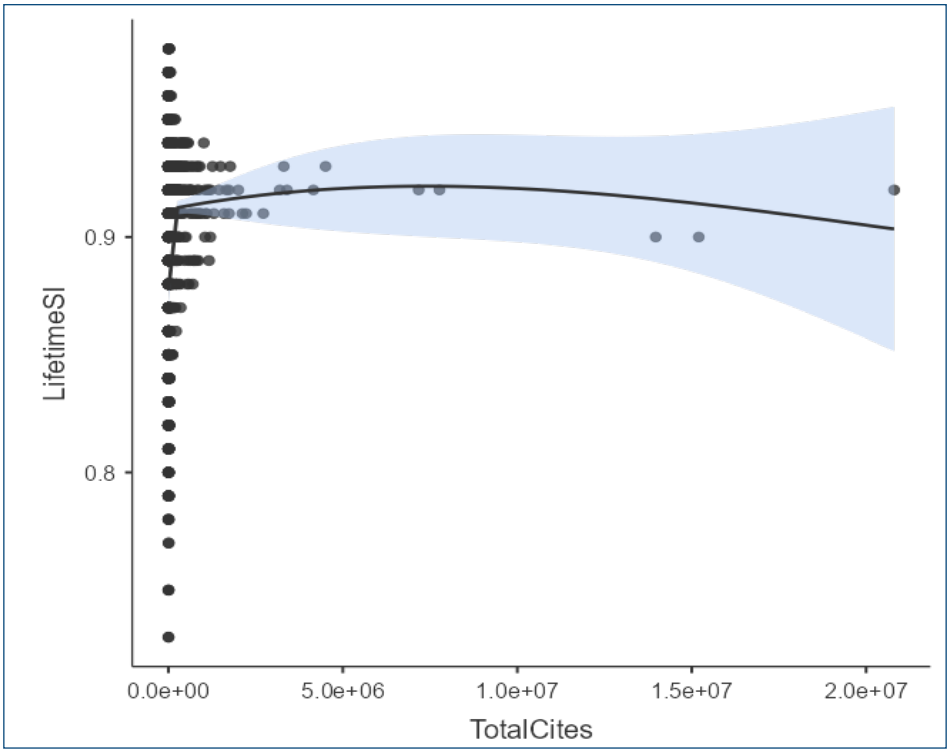


Fig. 3: Scatter Plot (Lifetime SI and Total Cites)

Pearson's correlation test was conducted between Lifetime SI (LSI) and Total Cites (TC) to create the correlation matrix. The Pearson correlation coefficient range from -1 to +1, with -1 indicating a perfect negative correlation and +1 indicating a perfect positive correlation. Correlation analysis was done to ascertain the strength and direction of the association between Lifetime SI and Total Cites.

The results of the correlation analysis are given in Table 5, which indicates there is a weak positive correlation (r=0.03) between Lifetime SI and Total Cites variables. It means that as the Lifetime SI of an organisation increases, there is a slight tendency for its total citations to increase as well, but the correlation is not strong enough to be statistically significant at the 0.05 level. However, the p-value (0.073>α) suggests that this correlation is not statistically significant at the 5% level of significance, meaning that it may have occurred by chance. The 95% confidence interval indicates that the true correlation coefficient lies between -0.003 and 0.063. Therefore, there is not enough evidence to conclude that there is a significant correlation between Lifetime SI and Total Cites variables. Hence, the null hypothesis ($H_0$1= "There is no significant relationship between "Lifetime SI and Total Cites" fails to be rejected, indicating the test has identified a weak positive relationship between the two phenomena Lifetime SI (LSI) and Total Cites (TC).

Table 6: Correlation Analysis between Total Cites and Contrasting

Correlation Matrix			
		Total Cites (TC)	*Contrasting (C)*
Total Cites	Pearson's r	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	N	—	
Contrasting	Pearson's r	0.973***	—
	p-value	< .001	—
	95% CI Upper	0.975	—
	95% CI Lower	0.972	—
	N	5000	—

Note. * p < .05, ** p < .01, *** p < .001

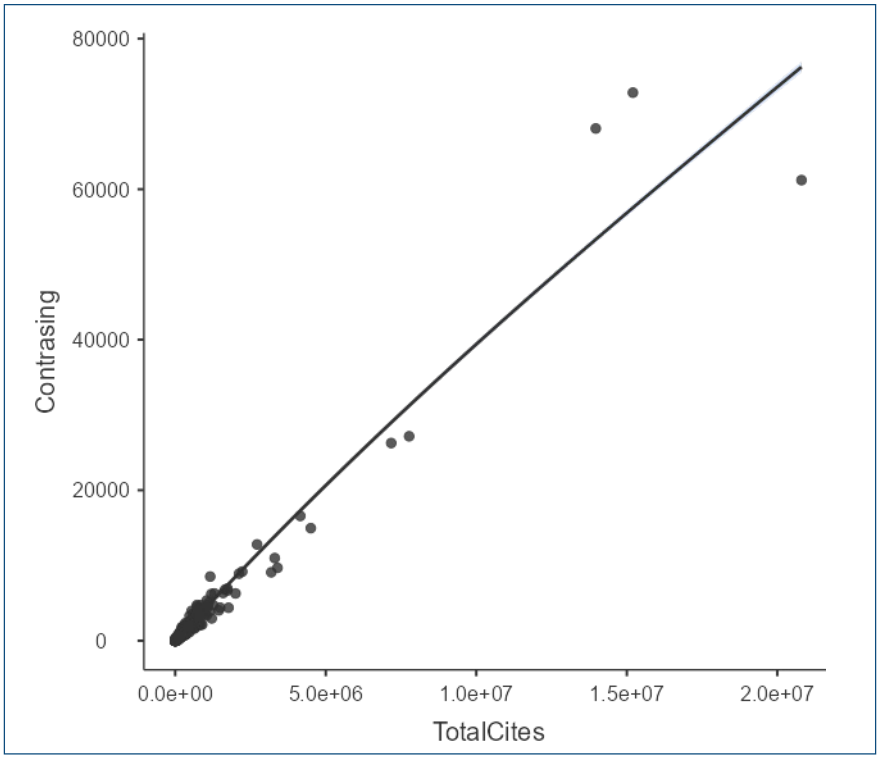


Fig. 4: Scatter Plot (Total Cites and Contrasting)

Pearson's correlation test was conducted between Total Cites (TC) and Contrasting (C) to create the correlation matrix. Correlation analysis was done to discover the strength and direction of the association between TC and C. The results of the correlation analysis are given in Table 6, which indicates a high degree of positive correlation (r= 0.973). A positive correlation between two variables means both variables move in the same direction. An increase in one variable leads to an increase in the other variable and vice versa. Thus, Total Cites TC and C exhibit a strong positive relationship, as depicted in Fig. 3. Furthermore, the p-value was less than the significance level or critical value ($p<\alpha=0.001$), indicating strong evidence against the null hypothesis. Hence, the null hypothesis ($H_0 2$= "There is no significant relationship exists between total citation and contrasting citations") stands rejected, indicating the test has identified a meaningful relationship between the two phenomena (TC and C).

Table 7: Correlation Analysis between Total Cites and Supporting

Correlation Matrix				
		Total Cites (TC)		*Supporting (S)*
Total Cites	Pearson's r	—		
	p-value	—		
	95% CI Upper	—		
	95% CI Lower	—		
	N	—		
Supporting	Pearson's r	0.988	***	—
	p-value	<.001		—
	95% CI Upper	0.988		—
	95% CI Lower	0.987		—
	N	5000		—

Note. * p < .05, ** p < .01, *** p < .001

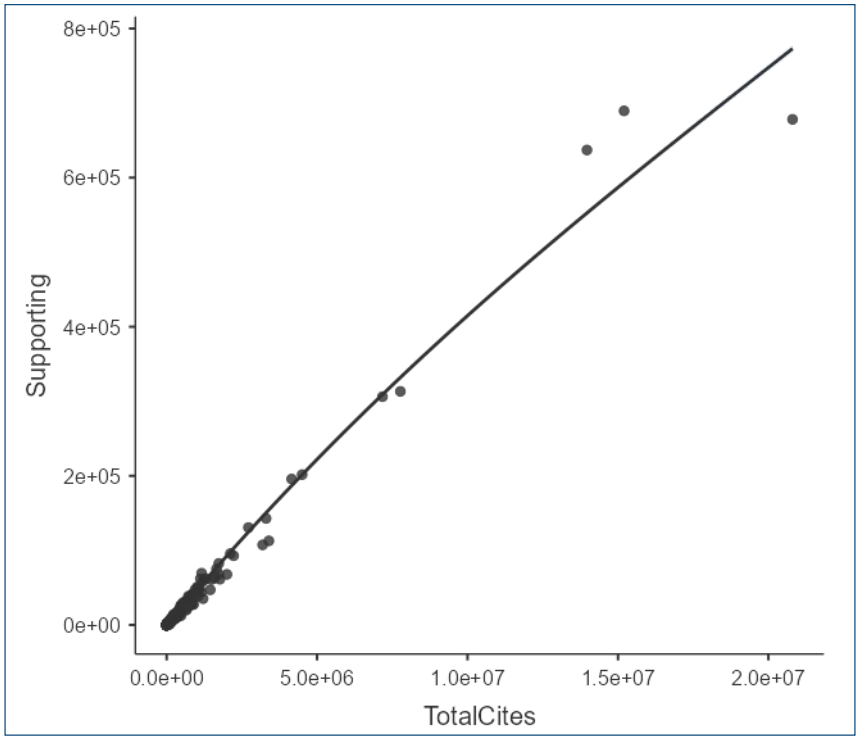


Fig. 5: Scatter Plot (Total Cites and Supporting)

To create the correlation matrix, Pearson's correlation test was conducted between TC and Supporting (S). Correlation analysis confirmed the strength and direction of the association between TC and S. The results of the correlation analysis are given in Table 7, indicating a high degree of positive correlation (r= 0.988). A positive correlation between two variables means both variables move in the same direction. An increase in one variable leads to an increase in the other variable and vice versa. Thus, TC and S exhibit a strong positive relationship, as depicted in Fig. 4. Furthermore, the p-value was less than the significance level or critical value ($p<\alpha=0.001$), indicating strong evidence against the null hypothesis. Hence, the null hypothesis ($H_0$3= "There is no significant correlation between Total Citations and Supporting Citations") stands rejected, indicating that the test has identified a consequential relationship between the two phenomena (TC and S).

Table 8: Correlation Analysis between Contrasting and Supporting

Correlation Matrix				
		Supporting (S)		*Contrasting (C)*
Supporting	Pearson's r	—		
	p-value	—		
	95% CI Upper	—		
	95% CI Lower	—		
	N	—		
Contrasting	Pearson's r	0.994	***	—
	p-value	<.001		—
	95% CI Upper	0.995		—
	95% CI Lower	0.994		—
	N	5000		—

Note. * p < .05, ** p < .01, *** p < .001

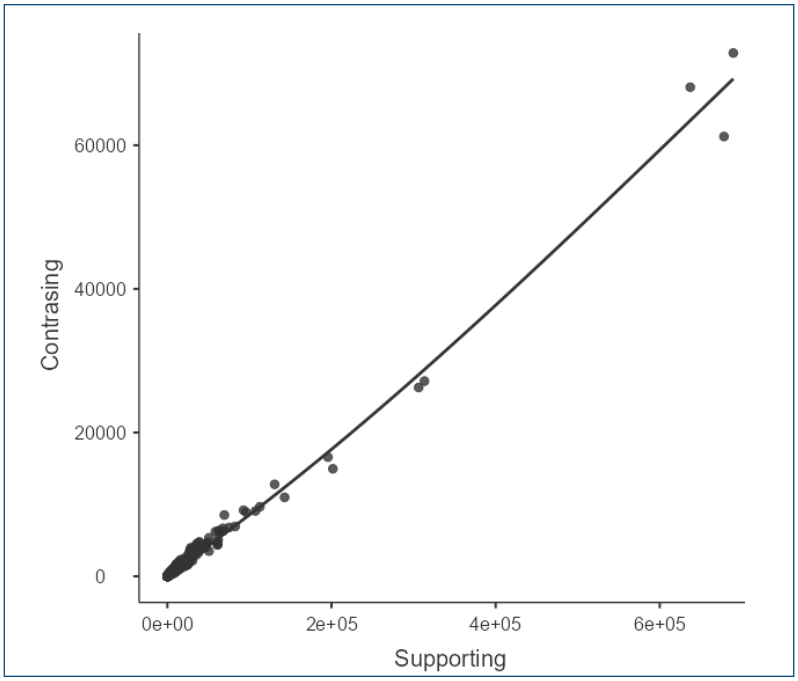


Fig. 6: Scatter Plot (Contrasting and Supporting)

Pearson's correlation test was conducted between C and S to create the correlation matrix. Correlation analysis was done to ascertain the strength and direction of the association between C and S. The results of the correlation analysis are given in Table 8, which indicates a high degree of positive correlation (r= 0.994). A positive correlation between two variables means both variables move in the same direction. An increase in one variable leads to an increase in the other variable and vice versa. Thus, C and S exhibit a strong positive relationship, as depicted in Fig. 5. Furthermore, the p-value was less than the significance level or critical value ($p<\alpha=0.001$), indicating strong evidence against the null hypothesis. Hence, the null hypothesis ($H_0$4= "There is no significant association exists between Contrasting and Supporting Citations") stands rejected, indicating the test has identified a significant relationship between the two phenomena (C and S).

Discussion

Contrasting research plays a vital role in scientific research by providing alternative perspectives, theories and findings that challenge existing ideas and assumptions. This research aims to identify and highlight differences and contradictions among studies, theories, or methods and to analyse the reasons for these differences. Contrasting research can help identify areas where previous research needs to be completed, biased or based on flawed assumptions and can offer insights and new directions for future research. By examining differences and contrasts between studies, researchers can identify limitations and areas where further investigation is needed, leading to more robust and comprehensive scientific findings.

Funding institutions in scite are essential for advancing scientific knowledge and innovation. They provide financial support for scientific research and development, which is critical for making significant discoveries and advancements in various fields of science. Scite offers several benefits for academic institutions and libraries. It helps improve research outcomes across the organisation by providing a better way to discover, contextualise and understand research. It has partnered with leading universities to help researchers find and evaluate literature more effectively. Scite helps with literature reviews and critical analysis by presenting what is said about a paper right away, saving time and boosting the quality of work. Libraries play a crucial role in providing access to scientific information and resources, and they often rely

on funding institutions to acquire the materials they need. This study delves into the world of funding institutions, examining their dashboards to find research productivity, compare rankings and track changes to their yearly scite index. The study also counts the number of supporting, mentioning and contributing factors to their research. By analysing these metrics, the study aims to provide a complete picture of the research productivity and how many citations were received for the funded publications. In examining the status of citation statements about the research productivity of funding Institutions using Scite, it was found that the "National Institutes of Health" has the highest contrasting citation, i.e., 72.8K, with a total citation of 480K and the "National Natural Science Foundation of China" has the highest Publication of 2.0M. It was also found that the disciplines of science and technology received the maximum funds from funding institutions, whereas fields like physics, law, innovation, environment and food received fewer funds.

The National Natural Science Foundation of China, has the highest number of retractions (1386) and corrections (6058) among the listed institutions and also has an increased number of withdrawals (314) and expressions of concern (49). This indicates that the higher the publications, the greater the number of retractions, corrections and withdrawals. The Errata category has the most articles, followed by corrections, and the Expressions of Concern category has the minor articles.

A study was conducted with four hypotheses involving Lifetime SI, TC, C and S. The correlation test rejected the null hypotheses and found a strong positive correlation between TC and C, TC and S and C and S. It can be observed that Lifetime SI and TC of the funding institutions exhibit a weak positive correlation (r=0.030). However, the weak correlation also indicates that this relationship is not very strong. TC and C show a 0.973 correlation, TC and S are correlated by 0.988, and C and S also show a strong positive correlation of 0.994. As a result, there was a significant association between the parameters TC and C, TC and S and C and S. Overall, the study suggests that TC, C and S are important variables that are strongly correlated with each other.

Conclusion

A citation analysis counts the number of times a scientific article is cited, and citation indexes like Scopus, and WoS, do not measure authors' intent to cite an article. Hence, all types of citations are counted as equal. A study has yet to be conducted that differentiates citation statement (citations) types, which helps assess the integrity of research articles. This study attempts to analyse the context of the citations with the help of an AI and DLM-based web tool, Scite. The platform facilitates a "smart citation index" that categorises citations according to context and is intuitive and powerful. It provides an in-depth look at not only the quantity but also the quality of citations. It evaluates scientific articles through Smart Citations–citations that display the context of the citation and describe whether the article provides supporting or disputing evidence. By presenting results in context, Scite presents more relevant and contextualised search results. Overall, Scite can help academic institutions and libraries better support research and innovation in their communities. The study findings attempted to provide a potential window on how to find retracted, withdrawn, errata, etc., articles within particular affiliations. There is a need to build an automated and comprehensive database of notices that detects retractions, withdrawals and other notices directly from the metadata of papers, such as their titles. Furthermore, future studies should examine different aggregate levels, such as journals and authors or the distribution of agreement and disagreement among individual articles.

References

Artificial Intelligence | Definition, Examples, Types, Applications, Companies, & Facts | Britannica. (2023). *Encyclopedia Britannica*. [Online] Retrieved February 19, 2023, from https://www.britannica.com/technology/artificial-intelligence

Brownlee, J. (2019). *What is deep learning*? [Online] Retrieved February 19, 2023, from https://machinelearningmastery.com/what-is-deep-learning/

Burns, E., & Brush, K. (2021). *Deep learning*. [Online] Retrieved February 19, 2023, from https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network

Chiong, W., & Haeusermann, T. (2019). Researcher integrity and contrasting obligations in interdisciplinary research. *J Empir Res Hum Res Ethics, 14*(5), 487-489. doi: https://doi.org/10.1177/1556264619853198a. PMID: 31779555; PMCID: PMC6980360.

Egger, A. E., & Carpi, A. (2010). *Scientific controversy.* [Online] Retrieved February 19, 2023, from https://www.visionlearning.com/en/library/Process-of-Science/49/Scientific-Controversy/181

Wager, E., Barbour, V., Yentis, S., & Kleinert, S. (2010). Retractions: Guidance from the committee on publication ethics (COPE). *International Journal of Polymer Analysis and Characterization, 15*(1), 2-6, doi:https://doi.org/10.1080/10236660903474522

Elsevier. (2016). Errata and corrigenda | Publishing ethics | Elsevier. [Online] Elsevier.com. Retrieved February 19, 2023, from https://www.elsevier.com/editors/perk/policy-and-best-practice-errata-And-corrigenda

Enago Academy. (2021). *Smart citations from scite - A new way to discover and understand research*. [Online] Retrieved February 19, 2023, from https://www.enago.com/academy/smart-citations-from-scite/

Futia, G., & Vetrò, A. (2020). On the integration of knowledge graphs into deep learning models for a more comprehensible AI—Three challenges for future research. *Information, 11*(2), p.122. doi:https://doi.org/10.3390/info11020122

Hayes, M. (2020). *Quantity over quality? Analyzing journal-level citations with scite*. [Online] Retrieved February 19, 2023, from https://masonrhayes.com/publication/scite_project/

Ibm.com.(2015). *What is deep learning?* [Online] Retrieved February 19, 2023, from https://www.ibm.com/in-en/topics/deep-learning#:~:text=Machine%20Learning%20Accelerator-what%20is%20deep%20learning%3F,from%20large%20amounts%20of%20data

Long, T. L. (2019). A history of citation styles. *Nurse Author & Editor, 29*(3), 1-6. doi:https://doi.org/10.1111/j.1750-4910.2019.tb00048.x

Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rodrigues, N. P., Grabitz, P., & Rife, S. C. (2021). Scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies, 2*(3), 882-898. doi:https://doi.org/10.1162/qss_a_00146

Plos.org. (2023). *PLOS Neglected Tropical Diseases.* [Online] Retrieved February 19, 2023, from https://journals.plos.org/plosntds/

The Process of Science. (2021, September 16). Retrieved from https://bio.libretexts.org/Courses/University_of_Pittsburgh/Environmental_Science_(Whittinghill)/02%3A_The_Process_of_Science

Prakash, K., & Sangam, S. L. (2013). *Improving access to open access journals: Abstracting, indexing and citation sources*. [Online] doi:https://doi.org/81-902079-3-8

R Core Team. (2021). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2022-01-01).

Retraction Watch. (2023). *Retraction watch.* [Online] Retrieved February 19, 2023, from https://retractionwatch.com/

Rife, S., Rosati, D., & Nicholson, J. (2021). s*cite: The next generation of citations scite: The next generation of citations.* [Online] Retrieved February 19, 2023, from https://digitalcommons.murraystate.edu/cgi/viewcontent.cgi?article=1096&context=faculty

Rosati, D. (2021). *How are journals cited? characterizing journal citations by type of citation*. [Online] doi:https://doi.org/10.48550/arXiv.2102.11043

Scite.ai. (2023). *About scite.* Retrieved from https://scite.ai/about

Study.com. (2023). *How observations can challenge existing scientific theory.* [Online] Retrieved February 19, 2023, from https://study.com/academy/lesson/how-observations-can-challenge-existing-scientific-theory.html

The Jamovi Project. (2022). *Jamovi.* (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org

Umd.edu. (2023). *Research guides: Bibliometrics and Altmetrics: Measuring the impact of knowledge:*

Promotion and tenure. [Online] Retrieved February 19, 2023, from https://lib.guides.umd.edu/c.php?g=327388&p=2196054

Uppala, A., Rosati, D., Nicholson, J. M., Mordaunt, M., Grabitz, P., & Rife, S. C. (2022). *Title detection: A novel approach to automatically finding retractions and other editorial notices in the scholarly literature*. [Online] doi:https://doi.org/10.48550/arXiv.2210.09553

Utexas.edu. (2023). *LibGuides: History: Introduction to Citations*. [Online] Retrieved February 19, 2023, from https://guides.lib.utexas.edu/c.php?g=708753&p=6235486