



The Impact of Google Scholar De-Indexing on Academic Visibility and Citation Performance: A Systematic Review

Delap Tonfold

Independent Researcher, UK

Corresponding Author: Delap Tonfold, E-mail: tonfold2111@gmail.com

ARTICLE INFORMATION	ABSTRACT
Submitted: April 02, 2026 Accepted: July 13, 2026 Published: August 26, 2026 Volume: 2 Issue: 1	Google Scholar has become a major platform for discovering scholarly literature, monitoring research impact, and supporting academic visibility through citation tracking. However, the de-indexing of scholarly publications from Google Scholar may reduce the discoverability of research outputs and potentially affect citation performance. This systematic review examines the impact of Google Scholar de-indexing on academic visibility and citation performance, with particular attention to the patterns, causes, and consequences of removal or loss of indexed scholarly records. The review synthesizes evidence from relevant studies and scholarly literature concerning Google Scholar indexing, academic discoverability, citation accumulation, metadata quality, technical accessibility, publisher practices, and research evaluation. The findings indicate that de-indexing can reduce the online discoverability of affected publications, interrupt citation tracking, create inconsistencies in author and institutional profiles, and limit the extent to which scholarly outputs contribute to visible citation-based indicators. The review further identifies technical problems, incomplete or inconsistent metadata, inaccessible full-text files, duplicate or malformed records, changes in publication platforms, and editorial or publishing-related issues as important factors that may contribute to indexing instability. However, the effects of de-indexing are not uniform, as citation performance is also influenced by publication quality, disciplinary practices, article age, collaboration networks, journal visibility, and availability through alternative scholarly databases and repositories. The study concludes that maintaining accurate metadata, technically accessible publications, stable web infrastructure, and consistent editorial practices is essential for sustaining Google Scholar visibility and reliable citation tracking. The review contributes to a better understanding of de-indexing as an important scholarly communication and research-impact issue and highlights the need for publishers, editors, researchers, and institutions to adopt proactive strategies for monitoring and maintaining academic discoverability.
KEYWORDS Google Scholar; de-indexing; academic visibility; citation performance; scholarly communication; research impact; indexing.	

1. Introduction

Academic visibility has become an increasingly important dimension of scholarly communication in the digital era. The discoverability of research through online academic search engines, bibliographic databases, institutional repositories, and publisher platforms influences how easily scholarly works can be located, read, shared, and cited. Among the platforms supporting this process, Google Scholar occupies a particularly prominent position because it provides broad access to scholarly literature across disciplines and indexes articles, conference papers, theses, books, preprints, and other academic materials (Van Hoboken, 2012). Researchers frequently use Google Scholar not only to locate literature but also to monitor citations, assess research influence, develop academic profiles, and identify potential collaborators. Consequently, changes in the indexing status of scholarly publications can have important implications for the visibility and perceived impact of researchers and journals.

Google Scholar indexing, however, is not guaranteed to be permanent. Scholarly documents may disappear from Google Scholar search results or become less discoverable because of editorial, technical, metadata, accessibility, duplication, or quality-related

factors. This phenomenon, commonly described as de-indexing, can occur at the level of individual publications, journal content, websites, or author profiles. De-indexing may therefore create uncertainty for authors and publishers, particularly when previously discoverable publications become difficult to locate through conventional Google Scholar searches (Cortegiani, 2020). Although the removal of a document from search results does not necessarily mean that the publication has been withdrawn, retracted, or deleted from the internet, reduced discoverability can nevertheless affect the pathways through which readers encounter and cite the work.

The relationship between discoverability and citation performance is particularly significant in contemporary scholarly communication. Citations depend on multiple factors, including research quality, relevance, disciplinary norms, publication venue, collaboration networks, accessibility, and the visibility of scholarly outputs. Digital discoverability represents an important enabling condition because researchers are more likely to encounter and potentially cite works that can be readily identified through academic search systems. When publications become difficult to retrieve, opportunities for exposure may decline, potentially contributing to slower citation accumulation (Teixeira da Silva, 2023). At the same time, citation counts should not be interpreted as a direct or exclusive measure of research quality, since citation practices vary substantially across disciplines, publication types, geographic regions, and research communities. Thus, the potential effect of Google Scholar de-indexing on citation performance requires careful examination rather than assuming a simple causal relationship.

The implications of de-indexing extend beyond individual researchers. For journals and publishers, persistent indexing contributes to the visibility of published content, facilitates the discovery of articles by prospective readers, and supports the development of citation networks. Reduced indexing coverage may consequently affect perceptions of journal reach and online presence. For institutions, research assessment exercises and academic performance monitoring may also rely, directly or indirectly, on bibliometric information obtained from scholarly search platforms (Savegnago de Mira, 2022). De-indexing can therefore introduce discrepancies between the actual existence of scholarly outputs and their visibility within a particular indexing environment. Such discrepancies are especially relevant for researchers and journals operating in developing scholarly ecosystems, where Google Scholar may provide an accessible avenue for discovering research that is not comprehensively represented in commercial citation databases.

Existing discussions of Google Scholar have primarily examined its coverage, citation-counting practices, profile management, indexing mechanisms, and limitations as a bibliometric tool. Research has also highlighted challenges associated with duplicate records, incorrect metadata, version aggregation, automated indexing, and differences between Google Scholar and other bibliographic databases (Mira, 2022). However, comparatively less attention has been devoted to systematically examining de-indexing as a distinct phenomenon and its potential consequences for academic visibility and citation performance. In particular, there remains a need to bring together evidence concerning the causes of de-indexing, the observable patterns through which it occurs, the mechanisms by which reduced discoverability may influence scholarly engagement, and the strategies that authors, editors, and publishers can use to diagnose and mitigate its effects.

Against this background, this systematic review examines the impact of Google Scholar de-indexing on academic visibility and citation performance. The review synthesizes available evidence on the nature and causes of de-indexing, its implications for the discoverability of scholarly publications, and its potential relationship with citation outcomes. It also considers the editorial, technical, metadata, accessibility, and quality-related factors that may contribute to changes in indexing status. By integrating these dimensions, the study seeks to clarify whether and under what circumstances de-indexing can reduce scholarly visibility and potentially affect citation trajectories (Pandita, 2022). The review further aims to identify gaps in the existing literature and provide evidence-informed implications for researchers, journal editors, publishers, librarians, and academic institutions seeking to maintain reliable and sustainable scholarly discoverability.

The study is guided by the broader recognition that academic impact is increasingly shaped by the interaction between research production and digital discoverability. Understanding Google Scholar de-indexing is therefore important not simply as a technical indexing issue but as a component of the wider scholarly communication ecosystem. A systematic examination of this phenomenon can contribute to more informed approaches to publication management, metadata quality, indexing compliance, research dissemination, and citation monitoring (Yoon, 2014). Ultimately, the review provides a structured foundation for understanding how changes in Google Scholar visibility may influence the pathways through which scholarly knowledge is discovered, accessed, disseminated, and subsequently incorporated into the academic literature.

2. Methodology

2.1 Research Design

This study employed a systematic literature review design to examine the impact of Google Scholar de-indexing on academic visibility and citation performance. A systematic review was considered appropriate because the study sought to identify,

evaluate, synthesize, and interpret existing evidence concerning the causes, patterns, consequences, and implications of the removal or loss of scholarly records from Google Scholar. Rather than treating de-indexing as an isolated technical problem, the review examined it as a multidimensional phenomenon involving editorial practices, technical accessibility, metadata quality, indexing policies, publication integrity, and changes in scholarly discoverability. The review followed a structured and transparent process involving literature identification, screening, eligibility assessment, data extraction, synthesis, and interpretation.

2.2 Review Questions

The review was guided by four principal questions: (1) What are the major factors associated with the de-indexing of scholarly publications from Google Scholar? (2) How does Google Scholar de-indexing affect the academic visibility and discoverability of researchers and publications? (3) What relationship exists between de-indexing and citation performance, including citation counts, citation accumulation, and researcher-level indicators? and (4) What editorial, technical, and ethical practices can publishers and researchers adopt to minimize the risk and consequences of de-indexing? These questions provided the analytical framework for identifying relevant studies and organizing the evidence obtained from the literature.

2.3 Search Strategy

A comprehensive literature search was conducted across major scholarly databases and academic information sources to identify publications relevant to Google Scholar indexing, de-indexing, academic visibility, and citation performance. The search covered peer-reviewed journal articles, conference papers, scholarly reviews, research reports, and authoritative technical or policy documents where these provided relevant evidence. Search terms were developed from the main concepts of the study and combined using Boolean operators. Core terms included "Google Scholar," "de-indexing," "deindexing," "index removal," "scholarly indexing," "academic visibility," "research visibility," "citation performance," "citation counts," "scholarly communication," "metadata," "crawler accessibility," and "academic publishing." Related terms were combined to broaden retrieval while maintaining relevance.

The search strategy was designed to capture literature addressing both direct evidence of Google Scholar de-indexing and closely related issues concerning scholarly discoverability and indexing reliability. Reference lists of relevant publications were also examined to identify additional studies that might not have been retrieved through database searches. Where appropriate, authoritative Google Scholar documentation and publisher or repository guidelines were consulted to contextualize findings concerning technical requirements and indexing practices.

2.4 Eligibility Criteria

Studies were considered eligible when they addressed at least one of the central concepts of the review, namely Google Scholar indexing or de-indexing, scholarly visibility, citation performance, indexing barriers, metadata quality, technical accessibility, publication integrity, or related dimensions of scholarly discoverability. Peer-reviewed research and academically credible publications were prioritized. Studies published in English were included to ensure consistency in screening and interpretation. Publications were excluded when they were unrelated to scholarly indexing, focused exclusively on general web-search optimization without an academic publishing context, lacked sufficient methodological or conceptual information, or duplicated evidence already represented in another publication.

The review also distinguished between direct and indirect evidence. Direct evidence included studies explicitly examining Google Scholar indexing or removal, whereas indirect evidence addressed factors such as metadata quality, repository accessibility, scholarly communication, and citation visibility that could help explain the consequences of indexing changes. This distinction enabled the review to synthesize evidence without incorrectly treating all literature as direct empirical evidence of Google Scholar de-indexing.

2.5 Study Selection and Screening

The retrieved records were subjected to a multi-stage screening process. Initially, duplicate records were identified and removed. Titles and abstracts were then screened against the predefined eligibility criteria. Publications that appeared potentially relevant were subjected to full-text assessment to determine whether they provided sufficient evidence to address the review questions. Studies that failed to meet the inclusion criteria at the full-text stage were excluded, with the principal reason for exclusion documented to promote consistency and transparency.

The selection process emphasized relevance to the specific relationship between indexing status, academic visibility, and citation performance. Particular attention was given to studies providing empirical evidence, documented examples, systematic analyses, or authoritative explanations of indexing mechanisms. The final body of literature therefore consisted of sources considered sufficiently relevant to support the review's thematic and analytical objectives.

2.6 Quality Assessment of the Included Literature

The methodological and conceptual quality of the included literature was assessed before synthesis. Consideration was given to the clarity of research objectives, appropriateness of research design, transparency of methods, adequacy of data sources, reliability of measurements, and consistency between reported evidence and conclusions. For conceptual and policy-oriented publications, emphasis was placed on the authority of the source, clarity of the evidence presented, and relevance to scholarly indexing and communication.

Greater interpretive weight was assigned to studies with clearly documented methods and stronger empirical evidence. Sources that provided useful contextual or technical information but lacked robust empirical designs were used primarily to support interpretation rather than to establish causal relationships. This approach was particularly important because the literature specifically addressing Google Scholar de-indexing and its effects on citation performance remains comparatively limited.

2.7 Data Extraction

A structured data-extraction framework was used to organize information from the selected publications. Extracted information included author and publication year, study objectives, geographical or disciplinary context where applicable, research design, data sources, indexing-related factors examined, reported effects on academic visibility, reported effects on citation performance, and major conclusions. Additional information concerning technical accessibility, metadata, publication quality, editorial practices, repository policies, and ethical concerns was extracted when relevant.

The extracted evidence was subsequently grouped according to the major dimensions of the review. These included patterns and causes of Google Scholar de-indexing, technical and editorial factors affecting indexing, consequences for academic visibility, consequences for citation performance, and strategies for prevention and mitigation. This categorization facilitated comparison across studies and supported the development of an integrated interpretation of the evidence.

2.8 Data Synthesis and Analysis

Because the included literature was expected to vary substantially in research design, objectives, data sources, and outcome measures, a narrative thematic synthesis was employed. Quantitative findings, where available, were interpreted alongside qualitative and conceptual evidence rather than statistically pooled. The analysis involved identifying recurring findings, comparing areas of agreement and disagreement, examining relationships between indexing status and scholarly outcomes, and organizing the evidence into broader explanatory themes.

The synthesis focused particularly on the pathway through which de-indexing may influence scholarly communication. The review considered whether reduced indexing and discoverability could limit opportunities for researchers to encounter, read, cite, and disseminate publications, while also recognizing that citation performance is influenced by numerous factors beyond Google Scholar visibility. Accordingly, the review did not assume a direct causal relationship between de-indexing and citation decline unless supported by the evidence. Instead, potential effects were interpreted in terms of visibility, discoverability, accessibility, citation opportunity, and researcher-level scholarly metrics.

2.9 Analysis of Academic Visibility and Citation Performance

Academic visibility was assessed conceptually through indicators such as discoverability of publications, accessibility through scholarly search systems, exposure to researchers, inclusion in citation networks, and opportunities for scholarly dissemination. Citation performance was examined through reported citation counts, citation accumulation, citation trends, and related bibliometric indicators where these were available in the reviewed literature. The analysis recognized that Google Scholar is only one component of the broader scholarly communication ecosystem and that citation outcomes may also depend on journal reputation, publication quality, field, publication age, collaboration patterns, accessibility, and dissemination practices.

The review therefore treated de-indexing as a potential disruption to the discovery pathway rather than as an automatic cause of reduced citations. This analytical distinction was important for avoiding overstatement of the available evidence and for identifying areas where future empirical research is needed.

2.10 Ethical Considerations

As the study relied exclusively on previously published literature and publicly available scholarly or technical information, it did not involve direct interaction with human participants or the collection of personally identifiable participant data. Nevertheless, ethical principles concerning accurate representation of evidence, appropriate attribution, and avoidance of unsupported claims were observed throughout the review process. Findings were synthesized without manipulating evidence to support predetermined conclusions. Particular care was also taken to distinguish documented indexing practices from interpretations or hypotheses concerning the reasons for individual cases of de-indexing.

2.11 Methodological Limitations

The methodology has several limitations. First, Google Scholar's indexing and ranking mechanisms are not completely transparent, which limits the ability of researchers to establish the precise reasons why individual publications or sources may become less visible or disappear from search results. Second, the literature specifically examining Google Scholar de-indexing and its direct effect on citation performance is relatively limited compared with the broader literature on scholarly indexing and bibliometrics. Third, differences in citation practices across disciplines and variations in the coverage of scholarly databases make direct comparison of citation outcomes difficult. Finally, restricting the review primarily to English-language literature may have excluded relevant evidence published in other languages. These limitations were considered when interpreting the findings and drawing conclusions from the synthesized evidence.

3. Findings and discussion

3.1 Patterns and Characteristics of Google Scholar De-Indexing

The systematic review indicates that Google Scholar de-indexing is best understood as a dynamic loss, reduction, or alteration of scholarly discoverability rather than a single, uniform event. The reviewed evidence shows that previously visible records may disappear from search results, citation counts may decline, author profiles may no longer display particular publications, or groups of publications may become less visible following changes in the websites on which they are hosted (Brock, 2016). Importantly, the literature does not establish a single standardized definition or measurement procedure for "de-indexing." Instead, related phenomena are frequently discussed in terms of coverage gaps, missing records, indexing failures, duplicate consolidation, inaccessible documents, and changes in citation counts. Google Scholar itself acknowledges that uninterrupted coverage of any particular source cannot be guaranteed and that records may be removed when websites become unavailable to its search robots or to most users.

The findings therefore suggest a distinction between temporary indexing fluctuations and persistent de-indexing. A temporary fluctuation may arise from delayed crawling, changes in website structure, or the time required for Google Scholar to process corrections and newly available content. Google reports that new papers are normally added several times a week, while corrections to already indexed papers can take substantially longer. Persistent de-indexing, by contrast, is more plausibly indicated when an article or group of articles remains absent after repeated title-based searches, particularly when the corresponding material remains available elsewhere and comparable records continue to be indexed (Marsh, 2024). This distinction is important because a temporary absence should not automatically be interpreted as evidence that a publication has been rejected, suppressed, or judged academically illegitimate.

The reviewed literature also shows that apparent de-indexing can sometimes represent record consolidation rather than genuine removal. Google Scholar groups different versions of the same work, such as preprints, repository copies, conference versions, and publisher versions, in order to improve search ranking and consolidate citations. Consequently, the disappearance of one URL or duplicate record does not necessarily mean that the underlying scholarly work has disappeared from Google Scholar (Pasquale, 2020). Previous empirical research similarly identified substantial deduplication within Google Scholar, demonstrating that changes in the number of visible records can partly reflect the platform's procedures for consolidating bibliographically related documents rather than outright exclusion.

3.1.1 Nature and Patterns of Google Scholar De-Indexing

Several patterns emerge from the evidence. First, de-indexing may occur at the individual article level, where a previously searchable article becomes difficult or impossible to retrieve. Second, it may affect an author's profile, particularly when individual publications disappear, are merged with other records, or are no longer correctly associated with the author's profile. Third, it can occur across a journal or publisher platform, especially where a technical or hosting problem affects multiple publications simultaneously (Bloch-Wehba, 2019). These levels should not be treated as equivalent: an absent article may represent an isolated metadata or accessibility problem, whereas simultaneous disappearance of numerous articles is more consistent with a source-level coverage problem.

A further pattern concerns reductions in citation counts. Google Scholar explains that citation counts may fall when either the cited article or papers citing it disappear from the web, become inaccessible to its robots, or are reformatted so that automated systems can no longer identify their bibliographic information and references. Thus, a declining citation count can be an indirect indicator of indexing disruption, but it is not by itself conclusive evidence of de-indexing. This finding is consistent with research demonstrating that Google Scholar data contain multiple forms of collection, publication, author, title, publisher, citation, and metadata errors (Romele, 2016). In one detailed examination, errors were identified in 279 of 281 examined references, although the authors cautioned that the unusually high rate was partly attributable to intensive manual checking.

The evidence further indicates that **coverage gaps are a persistent characteristic of academic search engines**, even though Google Scholar has historically achieved broad coverage. A comparative study using Crossref records found that Google Scholar failed to retrieve approximately 9.8% of the sampled publications, while the authors attributed differences in coverage partly to external factors such as web accessibility and robot-exclusion policies and partly to internal database requirements (Jørgensen, 2019). These findings reinforce the interpretation that not every missing record constitutes intentional de-indexing. Rather, absence from search results may emerge from the interaction between automated harvesting, website accessibility, bibliographic processing, and source-level requirements.

3.1.2 Causes and Triggers of De-Indexing

The findings identify **technical accessibility and crawler compatibility** as among the strongest documented factors associated with loss of coverage. Google Scholar's publisher guidance emphasizes that its systems must be able to identify, crawl, and process the full text and bibliographic information of scholarly documents. It also states that coverage gaps may result from technical problems in the automatic processing of publisher websites (Diab, 2018). Accordingly, inaccessible pages, broken links, restrictive server configurations, changes in website architecture, or barriers that prevent automated crawling can plausibly contribute to records becoming unavailable to Google Scholar.

Changes in URLs and hosting arrangements represent a related problem. When a journal migrates from one publishing platform to another, previously indexed pages may no longer resolve in the same way, while new pages may require fresh crawling and processing. Similarly, changes in PDF locations, redirects, page structures, or access arrangements may interfere with automated identification of bibliographic information (Polataiko, 2017). The evidence supports these mechanisms as technically plausible and, in some cases, directly documented; however, the systematic review does **not** justify claiming that every website migration causes de-indexing. Such relationships should therefore be described as associations unless a study directly demonstrates a causal sequence.

Metadata quality and duplicate records also emerged as important considerations. Google Scholar uses automatically extracted bibliographic information, including titles, authors, publication information, and citations, to identify and rank scholarly works. Google explicitly recommends authoritative metadata because it facilitates identification of citations and scholarly records. Research on Google Scholar errors has likewise identified problems involving titles, authors, publication years, publishers, citations, duplicates, missing URLs, reprints, translations, and inaccessible documents (Belategi, 2024). These findings indicate that apparent disappearance can sometimes result from incorrect parsing, record merging, or failure to associate a document with the expected bibliographic identity.

Publisher-platform changes constitute another potentially important trigger. When several articles disappear simultaneously after a platform migration, the pattern may indicate a source-level technical disruption rather than article-specific exclusion. Similarly, evidence that a journal's website becomes inaccessible to search robots can help explain why multiple publications lose visibility at approximately the same time (Batool, 2024). Google Scholar explicitly notes that when a source becomes unavailable to its robots or to a large proportion of users, it may remove the affected material until accessibility is restored.

The review found **less conclusive evidence concerning publication quality or scholarly legitimacy as direct causes of de-indexing**. Although academic search systems necessarily apply inclusion criteria and may exclude certain types of documents, the available evidence does not support the assumption that a missing article or journal has necessarily been removed because it is of poor quality, predatory, unethical, or academically illegitimate. Google Scholar's stated coverage criteria emphasize scholarly content, accessibility, crawlability, and bibliographic processing (Voss, 2015). Consequently, claims linking de-indexing directly to editorial quality should be treated cautiously unless supported by explicit evidence of a policy-based removal.

3.1.3 Variation Across Authors, Articles, and Journals

De-indexing does not appear to affect all scholarly entities uniformly. At the article level, individual records may be affected by missing metadata, inaccessible URLs, duplicate consolidation, or difficulties in processing a particular document. At the author level, the consequences can accumulate when several publications are affected simultaneously (Gordon, 2019). Researchers who rely extensively on Google Scholar for maintaining publication lists and monitoring citations may therefore experience a more noticeable effect than authors whose publications remain consistently available across multiple indexing and repository systems.

Variation also occurs according to **publication type and disciplinary context**. Google Scholar indexes journal articles, conference papers, theses, dissertations, books, preprints, technical reports, and other scholarly materials, but its coverage is not necessarily identical across these categories. Earlier comparative research found particularly broad Google Scholar coverage across scholarly journals, while other studies have emphasized the platform's distinctive inclusion of books, conference papers, non-US journals, and other materials that may be less comprehensively represented in traditional citation indexes (Cofone, 2020).

This breadth makes Google Scholar particularly valuable for researchers working in disciplines and regions where conventional indexing services may have narrower coverage, but it also means that indexing stability can vary across heterogeneous publication environments.

Publisher and hosting arrangements are similarly important. Publications hosted on stable, well-structured platforms with accessible full text and consistent metadata are more readily processed by automated indexing systems than materials hosted on unstable or technically restrictive websites. The review therefore identifies **digital hosting arrangements as an important contextual factor**, although the available literature does not permit a simple ranking of publisher platforms according to their risk of de-indexing (Aniwa, 2021). The strongest conclusion is that platform architecture, accessibility, metadata consistency, and continuity of web availability can influence coverage.

Publication age may also interact with these factors. Older articles may have undergone multiple platform migrations, URL changes, repository transfers, or metadata transformations, increasing the possibility that the currently accessible version differs substantially from the version originally crawled. Nevertheless, the evidence does not establish publication age itself as a direct cause of de-indexing (Islam, 2023). Rather, age may operate indirectly through cumulative changes in digital hosting and bibliographic representation.

3.2 Impact of De-Indexing on Academic Visibility and Discoverability

The findings indicate that the principal consequence of Google Scholar de-indexing is a **reduction in discoverability within an important academic search environment**. Google Scholar is designed to provide broad access to scholarly literature from publishers, repositories, universities, professional societies, and other academic sources. When a publication disappears from that environment, researchers who depend on Google Scholar may be less likely to encounter it during routine literature searches (Graham, 2014). The effect is therefore not necessarily equivalent to removal from the scholarly record; rather, it represents a reduction in the pathways through which the record can be discovered.

This distinction is particularly important because Google Scholar contributes not only to literature discovery but also to citation tracking and informal assessment of research visibility. Previous research has highlighted its broad citation coverage and its usefulness for disciplines and publication types that may be less well represented in traditional citation databases (Islam, 2023). De-indexing can therefore have consequences extending beyond simple search-result visibility, especially where researchers, institutions, or evaluators routinely consult Google Scholar.

3.2.1 Effects on Article and Author Discoverability

At the article level, de-indexing can reduce the probability that a publication will appear when researchers search by title, author, keywords, or subject. This can affect the article's opportunities to be read, cited, shared, or incorporated into subsequent research. The relationship should nevertheless be interpreted probabilistically rather than deterministically: researchers may still locate a missing article through Crossref, institutional repositories, library databases, publisher websites, DOI searches, or other scholarly search engines (Cuomo, 2023). The principal effect of de-indexing is therefore a **loss of one important discovery pathway**, not necessarily complete disappearance from scholarly communication.

For authors, the effect may be cumulative. When several publications are missing or inadequately represented, the researcher's Google Scholar profile may provide an incomplete picture of their output. Such incompleteness can affect the apparent visibility of the researcher and may contribute to changes in profile-level citation indicators. Google Scholar acknowledges that citation counts can decline when citing documents become unavailable to its robots or when bibliographic information can no longer be successfully processed (Islam, 2025). Consequently, de-indexing can influence both the discoverability of individual publications and the quantitative indicators derived from those publications.

The review also highlights the importance of distinguishing **actual declines in scholarly impact from declines in indexed visibility**. A reduction in Google Scholar citations does not necessarily mean that researchers have stopped citing the work. Instead, some citing documents may have become inaccessible, duplicated records may have been consolidated, or citation information may have become difficult for automated systems to parse. Studies of Google Scholar data errors demonstrate that these technical and bibliographic problems can materially affect citation datasets (Kurek, 2020). Therefore, citation declines following apparent de-indexing should be interpreted as changes in observable citation coverage rather than automatic evidence of declining intellectual influence.

3.2.2 Effects on Institutional and Journal Visibility

At the journal level, repeated or widespread indexing problems may reduce the visibility of the journal's articles in general scholarly searches. This can affect how easily prospective authors identify the journal, how readers encounter its publications,

and how reviewers and researchers become familiar with its content (Islam, 2025). The potential effect is particularly relevant for newer, regional, or specialized journals that depend heavily on open scholarly search services for international discovery.

For academic institutions and departments, the cumulative effect may be similarly significant. If publications produced by a university are incompletely represented in Google Scholar, the institution's research output may appear less visible through a commonly used discovery and citation platform. However, the review found insufficient evidence to conclude that de-indexing alone directly causes reductions in institutional reputation, author submissions, or collaboration (Amaya Camposeco, 2017). Such outcomes are plausible downstream consequences, but they depend on additional factors such as research quality, institutional reputation, journal prestige, networking, and the availability of alternative indexing systems.

The implications are particularly relevant because Google Scholar's broad coverage can benefit scholarly communities whose publications are not comprehensively represented in commercial citation indexes. Earlier research found that Google Scholar can provide broader coverage of books, conference papers, non-US journals, and other sources than conventional citation databases in some fields (Schwemer, 2021). Consequently, persistent indexing problems may have disproportionate consequences for journals and researchers whose visibility depends on freely accessible academic search infrastructure.

3.2.3 Implications for Scholarly Communication

At the broader level, the review shows that de-indexing can create **friction in the circulation of scholarly knowledge**. Scholarly communication depends on researchers being able to identify existing research, trace references, locate relevant evidence, and connect findings across disciplinary and geographic boundaries. When records become less discoverable, researchers may overlook relevant studies, particularly when searching through a single dominant academic discovery service (Van Hoboken, 2012). The problem is especially important for interdisciplinary research, where relevant publications may be distributed across different journals, repositories, disciplines, and publication formats.

Reduced discoverability may also contribute to **informational asymmetry**. Researchers with access to multiple databases, institutional subscriptions, library discovery systems, repositories, and bibliographic tools can potentially compensate for missing Google Scholar records more easily than researchers who depend primarily on free academic search services. This is significant because Google Scholar's free accessibility has been described as contributing to the democratization of citation analysis and providing broad access to citation information irrespective of institutional financial resources (Cortegiani, 2020). Thus, instability in its coverage may have a greater practical effect on researchers with fewer alternative discovery resources.

The findings also demonstrate why Google Scholar should not be treated as a **complete or infallible representation of the scholarly record**. Evidence of coverage gaps, duplicate records, missing URLs, inaccessible documents, metadata errors, and other bibliographic inconsistencies shows that search visibility is partly determined by technical and algorithmic processes (Teixeira da Silva, 2023). Consequently, researchers and institutions should interpret apparent de-indexing cautiously and, where possible, verify publication status through persistent identifiers, publisher pages, institutional repositories, and other bibliographic sources.

Overall, the systematic review indicates that Google Scholar de-indexing is best conceptualized as a **multidimensional visibility problem arising from the interaction of technical accessibility, metadata quality, record processing, platform changes, and source-level coverage mechanisms**. The evidence does not support treating every missing record as intentional removal or as a judgment of scholarly quality. Rather, de-indexing can range from temporary coverage fluctuations and duplicate consolidation to persistent loss of visibility caused by inaccessible or technically altered sources (Savegnago de Mira, 2022). Its principal consequence is a reduction in discoverability and potentially observable citation performance, with effects that may accumulate at the article, author, journal, and institutional levels. These findings reinforce the importance of maintaining technically accessible publication platforms, accurate metadata, stable URLs and persistent identifiers, while also using multiple discovery and preservation channels to reduce dependence on any single indexing service.

3.3 Impact on Citation Performance and Research Metrics

3.3.1 Changes in Citation Accumulation

The reviewed literature indicates that Google Scholar visibility is an important component of the citation process because discoverability determines whether potential readers and citing authors can readily locate a publication. Consequently, the loss or reduction of indexing can plausibly reduce opportunities for citation, particularly for newer or less-established publications that depend heavily on search-engine discovery. However, the evidence does not support treating de-indexing as an automatic or isolated cause of declining citation counts (Mira, 2022). Citation accumulation is a cumulative process influenced by research quality, topical relevance, disciplinary citation practices, publication age, collaboration networks, accessibility, and dissemination

through other scholarly channels. The appropriate interpretation is therefore that de-indexing may constrain one important pathway through which citations are generated rather than directly determining citation performance.

This interpretation is consistent with comparative studies showing that Google Scholar has exceptionally broad citation coverage. Martín-Martín et al. found that Google Scholar identified approximately 93–96% of citations to their sampled highly cited documents, compared with substantially lower coverage in Scopus and Web of Science. Google Scholar also identified many citations originating from theses, books, conference papers, and other non-journal materials. Therefore, when a legitimate publication disappears from Google Scholar, it may lose access to a particularly broad discovery environment (Pandita, 2022). For example, an article that remains available on a publisher's website but is no longer retrieved in Google Scholar searches may still be cited by researchers who encounter it through institutional repositories, reference lists, disciplinary databases, or direct publisher searches; nevertheless, its potential audience may be smaller than it would have been with continued indexing. The review consequently identifies reduced discoverability as a plausible mechanism for slower citation accumulation, while emphasizing that a direct causal relationship requires longitudinal evidence that controls for other determinants of citation growth.

3.3.2 Effects on h-Index and Other Author-Level Metrics

De-indexing can also affect author-level indicators when the omitted publications or their citing references contribute to the calculation of citation counts and the h-index. The effect is particularly important when an omitted article is close to the threshold required to increase an author's h-index (Yoon, 2014). For instance, if an author has several publications that each meet or nearly meet the relevant citation threshold, the disappearance of one highly cited record can alter the calculated h-index even though the underlying scholarly work and its citations have not actually disappeared. Thus, a change in a Google Scholar profile should not automatically be interpreted as a change in the author's actual research impact.

Previous comparative research demonstrates that citation and h-index values vary substantially according to the database used. Brock (2016) found different h-index ratings for the same researchers across Google Scholar, Scopus, and Web of Science and concluded that more than one source should be considered when assessing research impact. Similarly, Marsh (2024) reported that Google Scholar produced, on average, higher publication, citation, and h-index values than Web of Science in their sample of soil researchers. These differences mean that a reduction in a Google Scholar h-index following suspected de-indexing should be understood as a change in the indicator generated by that database, not necessarily as evidence that the researcher's broader scholarly influence has declined. The findings therefore reinforce the importance of identifying the data source whenever h-index, citation count, or related indicators are reported.

3.3.3 Reliability and Comparability of Citation Indicators

The review identifies database coverage as a central limitation affecting the reliability and comparability of citation indicators. Google Scholar, Scopus, and Web of Science do not operate as interchangeable databases: they differ in coverage, document types, indexing approaches, metadata quality, duplicate handling, and treatment of different versions of scholarly works. Google Scholar's comparatively inclusive approach allows it to capture citations from sources that may not be included in selective citation indexes (Pasquale, 2020). A large comparative study found that Google Scholar identified nearly all citations found by Web of Science and Scopus while also identifying substantial numbers of additional citations.

These differences have two implications for interpreting de-indexing. First, the same researcher or article may have different citation counts depending on the database consulted. Second, an apparent loss of citations may reflect database coverage or record-management differences rather than a genuine change in scholarly influence. The literature therefore supports making comparisons within the same database and under consistent measurement conditions. Bloch-Wehba (2019) specifically recommended that comparisons of researchers be conducted within a specified database because h-index rankings differ across systems. At the same time, Google Scholar's publisher guidance indicates that multiple versions of a work may be grouped so that citations to different versions can contribute to the scholarly record. This feature can increase citation coverage but also demonstrates why citation indicators require careful interpretation. Overall, the findings suggest that citation metrics are useful indicators of visibility and scholarly attention but should not be treated as perfectly stable or directly comparable measures across databases.

3.4 Technical, Editorial, and Ethical Dimensions of De-Indexing

3.4.1 Technical and Website-Related Factors

The review shows that technical accessibility is fundamental to continued scholarly discovery. Google Scholar relies on automated systems to identify and process scholarly documents, meaning that publications must be presented in ways that allow automated systems to retrieve and interpret them. Google Scholar's publisher guidance explicitly emphasizes technical and web-accessibility requirements for scholarly publishers and explains that the service indexes peer-reviewed papers, theses,

preprints, abstracts, and technical reports (Romele, 2016). Consequently, inaccessible full texts, crawler restrictions, inappropriate metadata, unstable URLs, poorly structured pages, JavaScript-dependent content, duplicate or confusing versions, and poorly formatted documents can interfere with discovery and indexing.

Technical problems may therefore produce apparent de-indexing without implying any substantive judgment about the quality of the underlying research. For example, an article may remain legitimately published while changes to the journal's website make the PDF difficult for automated systems to retrieve or the bibliographic information difficult to interpret. Similarly, migration from one publishing platform to another may alter URLs or page structures and temporarily disrupt discoverability (Jørgensen, 2019). The findings consequently support treating technical diagnosis as an essential first stage before attributing loss of visibility to editorial or quality-related causes. The distinction is important because technical non-discoverability and deliberate exclusion for publication-quality reasons have very different implications for authors and publishers.

3.4.2 Editorial and Publication-Quality Factors

The reviewed evidence indicates that editorial characteristics must be considered separately from technical indexing problems. Scholarly databases operate according to their own coverage and inclusion mechanisms, and the existence of an indexing problem does not by itself demonstrate that a journal has poor editorial standards (Diab, 2018). Accordingly, claims that a publication was removed because of inadequate peer review, questionable practices, retractions, or editorial deficiencies should be made only when there is documentary evidence supporting that conclusion.

At the same time, responsible editorial practices remain important to sustainable scholarly visibility. Transparent journal policies, clearly identified editorial responsibilities, consistent bibliographic information, credible peer-review procedures, correction and retraction mechanisms, and accurate publication records contribute to the integrity and usability of scholarly communication. Google Scholar describes its publisher-support framework as encompassing policies and technical information for scholarly publishers, illustrating that indexing is connected to the way scholarly content is presented and maintained online (Polataiko, 2017). The findings therefore suggest that publishers should avoid conflating indexing status with journal quality while simultaneously recognizing that persistent editorial weaknesses can damage trust, discoverability, and the long-term stability of a journal's scholarly record.

3.4.3 Ethical and Research-Integrity Considerations

De-indexing has significant ethical implications because visibility is part of effective scholarly communication. Authors reasonably expect legitimate publications to remain discoverable, readers need reliable access to the scholarly record, and institutions depend on bibliographic information for research assessment and knowledge dissemination (Belategi, 2024). If legitimate research becomes difficult to locate because of unexplained indexing problems, the resulting loss of visibility may disproportionately affect early-career researchers, less-established journals, researchers from regions with fewer dissemination resources, and publications in fields that depend strongly on broad web discovery.

At the same time, maintaining visibility cannot be separated from research integrity. Scholarly databases and publishers have a legitimate interest in ensuring that unreliable, fraudulent, retracted, or otherwise problematic material is appropriately identified and handled. The ethical challenge is therefore to balance protection of the scholarly record with procedural fairness and transparency. Where indexing changes occur, stakeholders should distinguish between technical failures, routine database changes, legitimate corrections to the scholarly record, and quality-related interventions (Batool, 2024). Transparent communication is particularly important because unexplained disappearance can lead authors or readers to infer misconduct when the actual cause may be technical or administrative. The findings thus support a principle of proportionality: legitimate research should not lose visibility unnecessarily, while justified interventions concerning unreliable research should preserve the integrity and traceability of the scholarly record.

3.5 Strategies for Prevention, Monitoring, and Recovery

3.5.1 Monitoring and Early Detection of De-Indexing

The review indicates that early detection is preferable to discovering indexing problems only after citation performance or author-level metrics have changed substantially. Authors, editors, and publishers can monitor the number and identity of indexed publications, check whether previously discoverable articles remain retrievable, compare author-profile records with institutional publication lists, and investigate unexpected discrepancies between Google Scholar and other databases (Voss, 2015). Sudden changes in indexed records, missing publications, unusual citation-count differences, or disappearance of previously accessible versions should be treated as signals requiring verification rather than immediate evidence of formal de-indexing.

Monitoring should be systematic rather than based on a single search. Because search results can vary according to query formulation, author-name ambiguity, version grouping, and database updates, a suspected indexing problem should be confirmed using multiple searches and, where possible, comparison with the publisher's records and other bibliographic databases. This approach is consistent with evidence that citation databases have different coverage and produce different author-level metrics (Gordon, 2019). Regular monitoring can consequently distinguish an isolated search-result problem from a broader disappearance of records and can help identify technical problems before they have substantial consequences for scholarly visibility.

3.5.2 Technical and Editorial Measures for Maintaining Visibility

Maintaining technically accessible and consistently organized journal platforms emerged as a central preventive strategy. Publishers should provide stable article URLs, accessible full-text files, clear bibliographic metadata, descriptive titles and author information, consistent publication dates, and webpages that can be processed by automated scholarly systems. Journal platforms should also minimize unnecessary barriers to accessing scholarly documents and carefully manage changes to website architecture so that previously published articles remain discoverable (Cofone, 2020). These practices are particularly important because Google Scholar explicitly provides technical guidance to publishers concerning the accessibility and indexing of scholarly content.

Technical maintenance should be complemented by editorial quality assurance. Journals should preserve accurate publication histories, clearly communicate peer-review and editorial policies, maintain correction and retraction procedures, and ensure consistency between article webpages, PDFs, metadata, and indexing records. Google Scholar's handling of multiple versions also illustrates the importance of coherent version management, since different versions can be grouped to consolidate citations (Aniwa, 2021). The findings therefore suggest that indexing sustainability should be treated as an ongoing publishing responsibility rather than a one-time technical achievement.

3.5.3 Implications for Authors, Publishers, and Research Institutions

The findings assign complementary responsibilities to authors, publishers, editors, and research institutions. Authors should maintain accurate publication lists, verify that their legitimate publications remain discoverable, monitor inconsistencies across scholarly databases, and use multiple legitimate dissemination channels rather than depending exclusively on one search service. Publishers and editors should maintain technically accessible platforms, stable bibliographic records, transparent editorial policies, and reliable procedures for corrections and retractions (Islam, 2023). Institutions, meanwhile, should avoid evaluating researchers solely through Google Scholar citation counts or h-index values because different databases produce systematically different measurements. Comparative evidence demonstrates that Google Scholar generally provides broader citation coverage than Scopus and Web of Science, while the resulting metrics can consequently differ substantially.

A major gap identified by the review is the limited availability of longitudinal empirical studies that follow individual publications before and after confirmed de-indexing events. Existing comparative research provides strong evidence about differences in database coverage and citation indicators, but this is not equivalent to demonstrating that a specific de-indexing event caused a specific reduction in subsequent citations. Future research should therefore combine longitudinal citation tracking with confirmed indexing-status changes, control groups, publication age, discipline, article quality, collaboration, open-access status, and alternative dissemination routes (Graham, 2014). Such research would provide stronger evidence about the magnitude and duration of the citation consequences associated with de-indexing. Overall, the findings support a balanced interpretation: Google Scholar is an important component of scholarly visibility, but its indexing status should be monitored as one element of a broader research-dissemination and impact strategy rather than treated as a complete measure of scholarly influence.

4. Conclusion

The findings of this systematic review demonstrate that Google Scholar de-indexing can have important implications for academic visibility, discoverability, and citation performance, particularly for researchers and journals that rely heavily on the platform as a major channel for scholarly dissemination. The reviewed evidence indicates that de-indexing may arise from a combination of technical problems, changes in web accessibility, metadata inconsistencies, duplicate or malformed records, publisher and repository practices, and concerns related to scholarly quality or indexing compliance. Although removal from Google Scholar does not necessarily mean that research has lost its academic value, reduced visibility can limit opportunities for other researchers to discover, read, cite, and build upon affected publications. Consequently, de-indexing may indirectly contribute to slower citation accumulation and weaker research metrics, especially for early-career researchers, institutions, and journals with limited alternative indexing coverage. The review therefore highlights the need for publishers, editors, repositories, and researchers to maintain technically accessible and accurately structured scholarly records, monitor indexing status regularly, and diversify dissemination through reputable databases and repositories. Overall, Google Scholar de-indexing should be

understood not merely as a technical indexing issue but as a scholarly communication concern with potential consequences for the visibility, reach, and measurable impact of academic research.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers

Artificial Intelligence (AI) Use Disclosure: The authors declare that no artificial intelligence tools were used in the preparation of this manuscript.

References

- [1] Amaya Campoeseo, J. E. (2017). The right to be forgotten: a descriptive overview of the right to be forgotten.
- [2] Aniwa S.C. (2021). Impact of Supply Chain Integration on Operational Performance: Insights from Manufacturing Firms in Emerging Economies. *Iconic Research And Engineering Journals*, 5(4).
- [3] Batool, A., Naseem, M., & Toyama, K. (2024, May). Expanding concepts of non-consensual image-disclosure abuse: A study of NCIDA in Pakistan. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).
- [4] Belategi, M. A. (2024). The right to be forgotten concerning the criminal past: Developments in the case law of the European Court of Human Rights with particular reference to the anonymisation of digital press archives. *Oñati Socio-Legal Series*, 14(6), 1639-1669.
- [5] Bloch-Wehba, H. (2019). Global platform governance: private power in the shadow of the state. *SMu L. reV.*, 72, 27.
- [6] Brock, G. (2016). *The right to be forgotten: privacy and the media in the digital age*. Bloomsbury Publishing.
- [7] Cofone, I. (Ed.). (2020). *The right to be forgotten: A Canadian and comparative perspective*. Routledge.
- [8] Cortegiani, A., Ippolito, M., Ingoglia, G., Manca, A., Cugusi, L., Severin, A., ... & Giarratano, A. (2020). Citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS (t) copus Project. *F1000Research*, 9, 415.
- [9] Cuomo, M. T., & Foroudi, P. (Eds.). (2023). *Digital transformation and corporate branding: Opportunities and pitfalls for identity and reputation management*. Taylor & Francis.
- [10] Diab, R. (2018). Search Engines and Global Takedown Orders: Google v Equustek and the Future of Free Speech Online. *Osgoode Hall LJ*, 56, 231.
- [11] Gordon, J. C. (2019). The perils of vanguardism: Explaining radical cuts to unemployment insurance in Sweden. *Socio-Economic Review*, 17(4), 947-968.
- [12] Graham, C. (2014). The calculation of age. *Organization Studies*, 35(11), 1627-1653.
- [13] Islam, A., & Jantan, A. H. B. (2023). The mediation effect of affective organizational commitment on the relationship between HRM practices and turnover intention in the RMG industry. *Research Journal in Business and Economics*, 1(1), 24-38.
- [14] Islam, A., Jantan, A. H. B., Khalifa, G. S., Islam, A., Islam, B., & Hossain, A. (2023). Effects of Decision Making and Work-life Balance on Productivity of Female Employees in the RMG Industry of Bangladesh. The Mediating Role of Work Motivation. *Research Journal in Business and Economics*, 1(1), 48-59.
- [15] Islam, M. A., & Sinniah, S. (2025). Exploring customer relationship management factors, customer trust, and innovation capacity: A quantitative study on customer retention. *Accountancy Business and the Public Interest*, 41(10), 12-29.
- [16] Islam, M. A., Islam, M. A., Amin, M. B., Hossain, M. M., Hassan, M. S., Afrin, S., & Oláh, J. (2025). Enhancing academic's performance: Exploring the interaction of innovative work behavior, intrinsic motivation, and self-efficacy in public universities. *Social Sciences & Humanities Open*, 12, 102210.
- [17] Jørgensen, R. F. (2019). Rights talk: in the kingdom of online giants. *Human rights in the age of platforms*, 163.
- [18] Kurek, M. R., Poulin, B. A., McKenna, A. M., & Spencer, R. G. (2020). Deciphering dissolved organic matter: Ionization, dopant, and fragmentation insights via Fourier transform-ion cyclotron resonance mass spectrometry. *Environmental Science & Technology*, 54(24), 16249-16259.
- [19] Marsh, A., & Lerner, A. (2024). Privacy norms of transformative fandom: A case study of an activity-defined community. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), 1-29.
- [20] Mira, B. S. D., Oliveira, C. G. D., & Shintaku, M. (2022). [RETRACTED ARTICLE] Editorial path and the visibility of scientific publications: an exploratory case study on the journal Tchê Química. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, 20, e022011.
- [21] Pandita, R., & Singh, S. (2022). Journal packing density of library and information science research journals at the global level: a study. *Global Knowledge, Memory and Communication*, 71(8-9), 857-872.
- [22] Pasquale, F. (2020). Internet nondiscrimination principles revisited. *Brooklyn Law School, Legal Studies Paper*, (655).
- [23] Polataiko, M. (2017). Best We Forget: Expressive Freedom and the Right to be Forgotten. In *7th USENIX Workshop on Free and Open Communications on the Internet (FOCI 17)*.
- [24] Romele, A. (2016). Digital Traceability and the right to be forgotten. Ricoeurian perspectives. *Tropos. Journal of Hermeneutics and Philosophical Criticism*, 2(8).
- [25] Savegnago de Mira, B., Gomes de Oliveira, C., & Shintaku, M. (2022). Editorial path and the visibility of scientific publications: an exploratory case study on the journal Tchê Química. *RDBCI: Digital Journal of Library & Information Science/Revista Digital de Biblioteconomia e Ciência da Informação*, 20, 1.
- [26] Schwemer, S. F., Mahler, T., & Styri, H. (2021). Liability exemptions of non-hosting intermediaries: Sideshow in the Digital Services Act?. *Oslo Law Review*, 8(1), 4-29.
- [27] Teixeira da Silva, J. A., & Nazarovets, S. (2023). Can the principle of the 'right to be forgotten' be applied to academic publishing? Probe from the perspective of personal rights, archival science, open science and post-publication peer review. *Learned Publishing*, 36(4), 651-666.

- [28] Van Hoboken, J. (2012). *Search engine freedom: on the implications of the right to freedom of expression for the legal governance of Web search engines* (Vol. 432). Alphen aan de Rijn: Kluwer Law International.
- [29] Voss, W. G., & Castets-Renard, C. (2015). Proposal for an international taxonomy on the various forms of the right to be forgotten: A Study on the Convergence of Norms. *Colo. Tech. LJ*, 14, 281.
- [30] Yoon, Y. J. (2014). Legal Ramifications of "Black Hat" Search Engine Optimization. *BUJ Sci. & Tech. L.*, 20, 159.