Stout in Computer Science and Technology Studies (SCSTS)

DOI: https://doi.org/10.61424/gt483298
Journal Homepage: https://stoutjournals.org/index.php/SCSTS

Understanding Google Scholar De-Indexing: A Review of Editorial, Technical, and Ethical Factors

Gregory Mutuse Heha
Independent Researcher, India
Corresponding Author: Gregory Mutuse Heha, **E-mail**: gmtuse12@gmail.com

ARTICLE INFORMATION

Submitted: March 21, 2026
Accepted: July 15, 2026
Published: August 26, 2026
Volume: 2
Issue: 1

KEYWORDS

Academic journal, Peer reviewed journal, Journal metadata, Google Scholar, Journal indexing.

ABSTRACT

Google Scholar plays a central role in the discovery, visibility, and citation of scholarly literature, making the removal or disappearance of journals and articles from its index a significant concern for authors, editors, publishers, and research institutions. This review examines Google Scholar de-indexing by synthesizing evidence on the editorial, technical, and ethical factors that may contribute to reduced or discontinued indexing. The review identifies editorial factors such as inconsistent publication practices, inadequate bibliographic information, questionable peer-review processes, irregular journal operations, and concerns regarding scholarly integrity. Technical factors include website accessibility problems, poorly structured article metadata, inappropriate document formatting, indexing barriers, changes in URLs or hosting platforms, and inadequate machine-readable content. Ethical considerations encompass plagiarism, duplicate or manipulated publications, misleading editorial practices, fabricated content, citation manipulation, and other forms of questionable publishing behavior that may undermine the reliability of scholarly records. The review further emphasizes that apparent de-indexing should be distinguished from temporary indexing delays, incomplete coverage, search-ranking changes, or technical disruptions. Overall, the findings indicate that Google Scholar visibility depends on the interaction of sound editorial governance, technically accessible publishing infrastructure, and adherence to research and publication ethics. The review concludes that systematic monitoring, accurate metadata management, transparent editorial practices, ethical compliance, and timely technical remediation are essential for maintaining and restoring scholarly discoverability.

1. Introduction

Google Scholar has become one of the most widely used scholarly discovery platforms, providing researchers, students, academic institutions, and publishers with access to a broad range of journal articles, conference papers, theses, books, and other scholarly materials. Its extensive coverage and relatively simple search interface make it an important mechanism through which academic publications are discovered, accessed, cited, and incorporated into subsequent research. For journals and publishers, visibility in Google Scholar can therefore contribute substantially to article discoverability, citation accumulation, academic reputation, and the dissemination of research findings (Stainforth, 2022). Consequently, the unexpected disappearance of journal articles or substantial reductions in their visibility within Google Scholar can generate significant concern among editors and publishers.

Google Scholar de-indexing refers broadly to the removal, exclusion, or loss of previously discoverable scholarly records from Google Scholar's search environment. However, apparent de-indexing is not always straightforward to identify because changes in search results may arise from several different circumstances, including technical indexing problems, changes in website structure, crawling limitations, duplicate records, metadata inconsistencies, temporary search fluctuations, or changes in Google's indexing processes (Maulidar, 2026). A reduction in the number of discoverable articles should therefore not automatically be interpreted as a formal or permanent removal of a journal from Google Scholar. Distinguishing genuine de-indexing from

temporary or technical visibility problems is an important challenge for journal editors and publishers seeking to maintain reliable scholarly dissemination.

The causes of Google Scholar de-indexing can be considered from interconnected editorial, technical, and ethical perspectives. From an editorial perspective, journals may encounter problems associated with inconsistent publication practices, inadequate bibliographic metadata, irregular article presentation, questionable editorial standards, or insufficient transparency concerning peer review and publication processes. Technical factors may include inaccessible article pages, inappropriate website configurations, unstable URLs, poor metadata implementation, robots.txt restrictions, indexing barriers, duplicate or fragmented versions of articles, and changes in journal platforms or domains (Popov, 2018). At the same time, ethical and scholarly-integrity concerns may become relevant when journals exhibit practices that conflict with accepted standards of academic publishing. These dimensions demonstrate that indexing is not solely a technical process but is closely connected to the quality, transparency, accessibility, and credibility of scholarly communication.

The implications of de-indexing extend beyond the immediate loss of search visibility. Reduced discoverability can limit the ability of researchers to locate relevant publications, potentially affecting article readership, citation opportunities, author visibility, and the broader dissemination of research findings. For journals, persistent indexing problems may undermine their perceived credibility and affect their attractiveness to prospective authors, reviewers, and academic institutions (Teixeira da Silva, 2023). In addition, uncertainty surrounding the reasons for disappearing records can create difficulties for editors who must determine whether the problem originates from their publishing practices, technical infrastructure, or external changes in search-engine indexing. These challenges are particularly important for emerging and smaller journals that depend heavily on scholarly search platforms to reach international research communities.

Despite the importance of Google Scholar for scholarly discovery, the phenomenon of de-indexing remains relatively underexamined compared with broader research on academic indexing, bibliometric databases, and journal evaluation. Existing discussions frequently focus on technical discoverability or journal quality without sufficiently integrating editorial and ethical dimensions (Tavani, 2012). A comprehensive understanding therefore requires consideration of the multiple factors that may contribute to changes in Google Scholar visibility and the difficulties involved in diagnosing and responding to such changes. Such an approach can help distinguish technical indexing failures from problems associated with editorial practices or scholarly publishing integrity.

Against this background, this review examines Google Scholar de-indexing through the interconnected lenses of editorial practice, technical infrastructure, and publication ethics. It synthesizes existing knowledge on the nature and observable indicators of de-indexing, potential editorial and technical causes, ethical considerations, consequences for journals and authors, and approaches to diagnosis and recovery (Askar, 2025). By bringing these dimensions together, the study seeks to provide a structured understanding of why scholarly content may become less visible in Google Scholar and to identify practical considerations for editors and publishers seeking to maintain sustainable scholarly discoverability. The review ultimately contributes to the broader discussion of responsible journal management, digital scholarly communication, and the preservation of reliable access to academic knowledge.

2. Methodology

2.1 Research Design

This study adopted a narrative review design to examine Google Scholar de-indexing from editorial, technical, and ethical perspectives. A review approach was considered appropriate because de-indexing is a multidimensional scholarly communication issue that involves journal editorial practices, website and metadata configurations, indexing mechanisms, publication integrity, and ethical standards. Rather than collecting primary data from individual journals, the study synthesized existing scholarly literature, indexing guidelines, publisher documentation, and relevant technical and ethical discussions to develop an integrated understanding of why academic journals and articles may disappear from Google Scholar and what such disappearance means for scholarly publishing.

2.2 Review Scope and Focus

The review focused specifically on Google Scholar de-indexing of academic journals and scholarly articles. The analysis covered three interconnected dimensions: editorial, technical, and ethical factors. Editorial factors included publication quality, journal policies, editorial consistency, questionable publishing practices, and compliance with scholarly communication standards. Technical factors included website accessibility, metadata quality, crawling restrictions, robots.txt configurations, article discoverability, URL changes, duplicate versions, and other factors that can affect automated indexing. Ethical factors included research integrity, plagiarism, fraudulent or manipulated content, misleading journal practices, transparency, and broader

concerns surrounding the credibility and reliability of scholarly information. The review also considered the distinction between genuine de-indexing and temporary or apparent loss of visibility caused by technical or indexing changes.

2.3 Literature Search Strategy

Relevant literature was identified through searches of major scholarly databases and academic information sources, including Google Scholar, Scopus, Web of Science, Crossref, and other relevant scholarly databases and publisher resources where appropriate. Search terms were developed around combinations of keywords such as "Google Scholar de-indexing," "Google Scholar indexing," "journal de-indexing," "scholarly article indexing," "academic journal visibility," "Google Scholar inclusion," "metadata and indexing," "web crawling and scholarly publishing," "publication ethics," and "research integrity." Additional searches used combinations of these terms with concepts relating to technical accessibility, journal quality, editorial practices, and ethical publishing. Reference lists of relevant publications were also examined to identify additional studies and authoritative materials that could contribute to the review.

2.4 Inclusion and Exclusion Criteria

Sources were considered eligible when they directly or substantially addressed scholarly indexing, Google Scholar, academic journal discoverability, technical factors affecting scholarly web indexing, editorial quality, publication ethics, or research integrity. Peer-reviewed journal articles, scholarly books and book chapters, conference papers where relevant, institutional publications, indexing guidelines, publisher documentation, and authoritative scholarly communication resources were considered where they provided useful evidence. Preference was given to recent and authoritative sources while retaining older foundational publications when they provided important conceptual or technical explanations. Sources that were unrelated to academic publishing, lacked sufficient relevance to scholarly indexing, consisted primarily of unsupported opinions, or did not provide meaningful information for understanding de-indexing were excluded.

2.5 Data Extraction and Organization

Information from the selected sources was systematically extracted according to the major themes of the study. Relevant evidence was organized into categories corresponding to editorial factors, technical factors, ethical factors, indicators of de-indexing, consequences of reduced indexing visibility, diagnostic approaches, and possible recovery or corrective measures. Particular attention was given to evidence explaining the difference between formal removal from an indexing system and temporary non-discoverability resulting from crawling, metadata, website, or indexing problems. This thematic organization enabled findings from different sources to be compared and integrated rather than treated as isolated observations.

2.6 Data Analysis and Synthesis

A thematic synthesis approach was used to analyze the reviewed literature. Findings were compared across sources to identify recurring explanations, areas of agreement, and important distinctions concerning Google Scholar indexing and de-indexing. The analysis examined how editorial, technical, and ethical factors may operate independently or interactively. For example, a journal may experience reduced discoverability because of technical barriers even when its editorial practices remain appropriate, whereas concerns about publication integrity may raise editorial or ethical issues that have broader implications for indexing status. The synthesis therefore treated de-indexing as a phenomenon that requires careful diagnosis rather than assuming that every decline in search visibility represents formal removal by Google Scholar.

2.7 Reliability and Quality of the Review

To enhance the credibility of the review, greater weight was given to peer-reviewed research and authoritative sources produced by recognized scholarly communication, indexing, publishing, and research-integrity organizations. Information from multiple sources was compared where possible to reduce dependence on a single explanation of de-indexing. Particular care was taken to distinguish documented indexing requirements from assumptions or anecdotal claims about Google's indexing decisions. Because Google Scholar does not publicly disclose every individual indexing or removal decision, the study avoided presenting unverified explanations as definitive causes. Instead, the review identified evidence-supported factors that may contribute to de-indexing or reduced discoverability and emphasized the importance of verification before concluding that a journal has been formally de-indexed.

2.8 Ethical Considerations

As a literature-based review, the study did not involve human participants, personal data collection, experimentation, or direct intervention with journals. Nevertheless, ethical considerations were incorporated into the interpretation of evidence. The review avoided making unsupported allegations against specific journals, publishers, editors, or authors and distinguished between confirmed misconduct and potential indicators of publication-quality problems. Sources were interpreted in their appropriate context, and claims concerning unethical publishing, research misconduct, or indexing violations were treated cautiously. The study also recognized that loss of visibility in Google Scholar does not, by itself, establish misconduct or poor editorial quality.

2.9 Limitations of the Methodology

The review has several limitations. First, Google Scholar's indexing processes are not completely transparent, making it difficult to determine the precise cause of every instance of apparent de-indexing. Second, the literature specifically addressing Google Scholar de-indexing is relatively limited compared with research on scholarly indexing more broadly. Consequently, evidence concerning general academic indexing, web crawling, metadata, research integrity, and journal quality was also considered when directly relevant. Third, indexing status can change over time because of technical modifications, algorithmic changes, website restructuring, or updates to scholarly databases. The findings should therefore be understood as a synthesized interpretation of the available evidence rather than as a universal diagnostic model applicable to every case of Google Scholar de-indexing.

3. Findings and Discussion

3.1 Patterns and Characteristics of Google Scholar De-Indexing

3.1.1 Nature and Patterns of Google Scholar De-Indexing

The review indicates that Google Scholar de-indexing is best understood as a dynamic loss or reduction of scholarly discoverability rather than a single, clearly defined removal event. Google Scholar itself notes that it cannot guarantee uninterrupted coverage of a particular source and that records may disappear when websites become unavailable to its search robots or to users. It also acknowledges that gaps in coverage may result from technical problems affecting the automatic processing of websites (Bhattacharjee, 2023). Consequently, an article that previously appeared in Google Scholar may become temporarily unavailable without necessarily indicating a permanent journal-level sanction. This distinction is important because apparent de-indexing can result from changes in article URLs, crawling interruptions, server errors, metadata-processing problems, or changes in how different versions of the same work are grouped.

The evidence further suggests that de-indexing can occur at different levels of visibility. At the article level, a previously discoverable paper may no longer be retrieved when its exact title is searched. At a broader level, several articles from the same journal may simultaneously become difficult to locate, creating the appearance of journal-wide de-indexing. However, Google Scholar emphasizes that it indexes papers rather than journals, meaning that claims that a journal has been completely "removed" should be treated cautiously and verified through systematic searches of individual articles (Yoon, 2014). This finding supports a more nuanced interpretation of de-indexing in which complete disappearance, partial coverage, and temporary discoverability problems are analytically distinguished rather than treated as equivalent phenomena.

Another important pattern is the difference between disappearance and reclassification or consolidation. Google Scholar groups multiple versions of a work, such as preprints, conference papers, repository copies, and publisher versions, in order to improve search ranking and consolidate citations. Therefore, the disappearance of one URL or version does not necessarily mean that the underlying scholarly work has been removed from the index. Previous research on Google Scholar has similarly highlighted the importance of deduplication and version management in interpreting its coverage and search results (Pandita, 2022). For this reason, evidence of de-indexing should ideally be based on repeated title-level searches, author searches, and examination of alternative versions rather than on a single failed search.

3.1.2 Scope and Extent of De-Indexing

The reviewed evidence shows that indexing problems can range from isolated article-level failures to broader losses involving groups of publications. An individual article may fail to appear because its bibliographic information cannot be correctly extracted, because its PDF is not searchable, because the article is hosted on an inaccessible URL, or because the article does not meet technical requirements for indexing (Cortegiani, 2020). Google Scholar specifically identifies incorrect extraction of titles, authors, publication information, and references as a common source of indexing problems. Such problems can affect one article independently of the rest of a journal.

More extensive patterns emerge when the same technical or structural problem is repeated across a journal's website. For example, if article pages do not contain machine-readable bibliographic metadata, if multiple papers are placed on the same webpage or in a single PDF, or if article URLs cannot be reached through a crawlable browse interface, many publications may experience reduced visibility simultaneously (Khatri, 2023). Google Scholar recommends that each paper have its own unique URL and that bibliographic information such as title, author, publication date, journal title, volume, and issue be presented in a form that automated parsers can process. Thus, the scope of de-indexing can provide diagnostic information: an isolated disappearance may suggest an article-specific problem, whereas simultaneous problems across numerous articles may indicate a journal-level website, metadata, hosting, or editorial-system issue.

The evidence also cautions against interpreting a decline in the number of visible results as definitive proof of journal-level removal. Google Scholar states that the site: search operator is not a reliable measure of total coverage because it searches primary versions and provides an estimated result count rather than a complete inventory. It recommends checking a sample of

individual article titles instead (Mira, 2022). This finding has methodological implications for studies of de-indexing: the extent of the problem should be established using systematic article-level verification rather than relying exclusively on the apparent number of search results.

3.1.3 Temporal and Recurring Patterns

Temporal analysis indicates that apparent de-indexing may occur through sudden, gradual, or fluctuating patterns. A sudden decline may follow a server outage, URL restructuring, changes to robots.txt, migration to a new publishing platform, or other alterations that prevent Google Scholar's crawlers from accessing article pages. Google Scholar explicitly states that server errors, misconfiguration, slow responses, or crawler restrictions can cause some or all articles to drop from Google Scholar (Frosio, 2020). Such events may therefore produce an abrupt reduction in discoverability without representing a permanent editorial judgment about the journal.

Gradual changes may occur because Google Scholar continuously crawls and updates its index rather than maintaining a static database. New papers are normally added several times a week, while corrections and updates to existing records can take substantially longer; Google Scholar indicates that some updates may require six to nine months or longer. Consequently, short-term differences in indexed records should not automatically be classified as de-indexing (Savegnago de Mira, 2022). A journal undergoing website migration, metadata correction, or content restructuring may experience a period during which some articles remain visible while others temporarily disappear.

Recurring fluctuations may likewise reflect the interaction between website accessibility, automated crawling, metadata extraction, duplicate detection, and version grouping. Google Scholar notes that it periodically revisits websites to refresh content and that changes in website accessibility can influence coverage. The findings therefore support distinguishing persistent de-indexing from transient indexing instability (Singson, 2025). A credible diagnosis should document the date of disappearance, monitor the article over time, test alternative article versions, and examine whether other publications from the same source remain discoverable.

3.2 Editorial Factors Associated with De-Indexing

3.2.1 Journal Quality and Editorial Standards

The review identifies editorial quality and the credibility of scholarly content as important contextual factors in understanding indexing problems, although technical evidence should not be confused with a direct claim that Google Scholar automatically removes journals because of poor editorial quality. Google Scholar's published inclusion guidance focuses primarily on whether a website hosts scholarly material, makes abstracts or full text accessible, and presents content in a technically crawlable and machine-readable form (Ananny, 2016). Nevertheless, editorial weaknesses can indirectly contribute to indexing difficulties when they are reflected in inconsistent article structures, inadequate metadata, unclear publication information, or content that does not consistently resemble conventional scholarly publications.

A well-managed journal is therefore more likely to maintain the structural consistency necessary for reliable indexing. Clear article titles, accurate author information, publication dates, journal details, abstracts, references, and stable article URLs facilitate automated identification and processing. Google Scholar specifically requires sufficient bibliographic information for its parsers and warns that incorrect extraction can result in papers being omitted, incorrectly identified, or ranked poorly (Sookman, 2018). This means that editorial quality control has a technical dimension: accurate editorial records are not merely important for human readers but also for automated scholarly discovery systems.

The review further suggests that weaknesses in peer-review documentation, authorship verification, conflict-of-interest procedures, and publication ethics should be considered as credibility risks, particularly when they coincide with other irregularities. However, these concerns should not be presented as automatic causes of Google Scholar de-indexing unless direct evidence establishes such a relationship. Rather, they form part of a broader quality-assurance environment in which journals should maintain transparent editorial policies and consistent scholarly standards (Lyseiuk, 2025). This interpretation avoids overstating the available evidence while recognizing that sustainable discoverability depends on both credible publishing practices and technically accessible content.

3.2.2 Publication Practices and Journal Content

Publication practices were found to have a substantial potential influence on the stability and quality of Google Scholar coverage. Irregular article structures, inconsistent metadata, duplicate records, multiple papers placed on the same webpage, and unstable article URLs can interfere with automated discovery and identification. Google Scholar states that it cannot effectively index multiple abstracts on the same webpage or multiple papers contained in a single PDF, and it requires each

paper to have its own unique URL (O'Hara, 2016). Consequently, journals that publish content using inconsistent or technically unsuitable formats may experience incomplete coverage even when their articles are genuinely scholarly.

The findings also highlight the importance of metadata consistency. Google Scholar's technical guidelines recommend machine-readable metadata, including article title, authors, publication date, journal title, volume, issue, and page information. Incorrect metadata can cause articles to be misidentified or treated as duplicates. For example, if a journal or repository name is incorrectly interpreted as the article title, multiple records may become difficult for the automated system to distinguish (Marinett, 2024). This demonstrates how apparently editorial problems may actually operate through technical mechanisms: poor publication practices affect the accuracy with which automated systems interpret scholarly records.

Questionable or highly repetitive content should also be considered cautiously. The presence of duplicate articles, substantially overlapping records, or content that departs from recognizable scholarly formats can complicate indexing and record consolidation. Google Scholar's emphasis on scholarly articles and its treatment of multiple versions and duplicate records indicates that discoverability depends partly on the system's ability to identify distinct scholarly works accurately (Brock, 2016). Therefore, journals seeking stable visibility should maintain consistent article-level publication structures, accurate metadata, identifiable authorship, stable URLs, and clearly distinguishable scholarly records.

3.2.3 Editorial Integrity and Ethical Compliance

Editorial integrity emerged as a critical dimension of the broader de-indexing problem because ethical failures can undermine the reliability and credibility of scholarly records. Issues such as plagiarism, fabricated or manipulated data, inappropriate citation practices, compromised peer review, undisclosed conflicts of interest, and questionable authorship undermine the integrity of the scholarly communication process (Islam, 2026). Although Google Scholar's public documentation does not establish that each of these ethical violations independently triggers automatic de-indexing, such practices may coincide with broader problems in the reliability, structure, and management of journal content. The distinction is therefore important: ethical misconduct should not automatically be described as a direct Google Scholar de-indexing criterion.

The review nevertheless indicates that ethical compliance has important implications for discoverability because trustworthy scholarly communication depends on accurate and stable records. Inappropriate citation practices or manipulated bibliographic information can interfere with automated reference extraction and citation linking. Google Scholar explains that its systems automatically extract bibliographic information and references and that incorrect identification can lead to poor indexing, incorrect author or title information, or even exclusion of papers (Islam, 2026). Ethical and editorial integrity are therefore connected to indexing indirectly through the accuracy and reliability of the scholarly record.

Peer-review irregularities and questionable publication practices also have consequences beyond individual articles. If concerns arise about a group of papers, journals may need to issue corrections, expressions of concern, retractions, or other editorial notices. Such changes can alter the bibliographic relationships among articles and their versions, potentially affecting how records are displayed or grouped. At the same time, the findings emphasize that a reduction in Google Scholar visibility should not itself be interpreted as proof of misconduct (Pasquale, 2020). Google Scholar's documentation identifies technical accessibility, crawling, metadata extraction, and content-format requirements as major explanations for coverage problems. Accordingly, ethical concerns should be investigated independently through appropriate editorial and publishing-ethics procedures, while de-indexing should be diagnosed using verifiable evidence of changes in scholarly discoverability.

3.3 Technical Factors Affecting Google Scholar Indexing

3.3.1 Website Accessibility and Crawler Compatibility

The review indicates that website accessibility and crawler compatibility are among the most important technical conditions influencing the continued visibility of scholarly articles in Google Scholar. Google Scholar relies on automated crawlers to discover, retrieve, and process scholarly documents; therefore, articles may become difficult to index when journal websites are inaccessible, return server errors, impose inappropriate access restrictions, or otherwise prevent automated retrieval (Slane, 2018). Google Scholar's publisher guidance explicitly states that publishers should ensure that article pages and full texts can be successfully identified, crawled, and processed, while its inclusion guidelines require scholarly content to be accessible through a crawlable website. In particular, articles should not depend on login screens, software installation, interstitial pages, or other barriers that prevent users and crawlers from reaching the abstract or full text.

The findings further suggest that technical accessibility is not simply a condition for initial inclusion but also for continued visibility. Google Scholar reports that it reflects the content currently visible to its search robots; consequently, when an article or the documents citing it become unavailable to crawlers, citation counts and search visibility can decline. This provides an important explanation for apparent de-indexing: a journal may remain operational from the perspective of human visitors while

its articles are intermittently inaccessible to automated systems (Van Hoboken, 2012). For example, a journal website that loads normally in a browser but temporarily returns server errors, blocks crawler requests, or places article content behind an inaccessible JavaScript-based interface may experience incomplete indexing. General Google technical guidance similarly identifies crawler blocking, unsuccessful HTTP responses, and non-indexable content as fundamental barriers to indexing.

These findings are consistent with Zhao (2022) case study of large fluctuations in Google Scholar coverage, which demonstrated that documents could disappear from Google Scholar when they ceased to be available to crawlers and, in some instances, remain absent for months before recovery. Importantly, their study also found that some documents disappeared even though they remained available on the wider Web, suggesting that indexing disruptions may sometimes originate from problems in the indexing system rather than from permanent removal by publishers. Thus, website accessibility should be treated as a continuous technical responsibility rather than a one-time requirement for journal inclusion.

3.3.2 Metadata, Document Structure, and Article Identification

The review identifies accurate metadata and recognizable document structure as another major determinant of Google Scholar visibility. Google Scholar automatically extracts bibliographic information from scholarly documents, including titles, authors, publication dates, journal information, and references. Its technical guidelines recommend structured metadata such as citation_title, citation_author, citation_publication_date, journal title, volume, issue, page information, and links to the article PDF (Gregory, 2018). At minimum, the article title, at least the first author's full name, and publication year need to be identifiable for effective processing.

The evidence suggests that metadata problems can produce several different outcomes rather than simple inclusion or exclusion. Incorrect or incomplete metadata may cause an article to be indexed under an incorrect title, author, publication year, or journal, making it difficult to retrieve through conventional searches. Google Scholar's guidelines note that bibliographic information is automatically extracted and that inaccurate data may prevent the system from matching an article with references that cite it. Similarly, studies examining Google Scholar data have identified errors involving titles, authors, publication years, publication information, missing URLs, inaccessible documents, and duplicate records. A detailed investigation by Aniwa (2021) found substantial errors across different components of Google Scholar records, demonstrating that automated scholarly indexing is vulnerable to data-quality problems.

Document structure is particularly important for journal publishers because an article may be technically available but difficult for automated systems to interpret. Google Scholar recommends that each article have a unique URL and that individual articles and abstracts be presented as separate HTML or PDF documents. PDF files should contain searchable text rather than merely scanned images, and article titles and author information should be clearly distinguishable from journal or website information (Islam, 2025). Consequently, poorly structured PDFs, image-only documents, multiple articles combined into a single file, inconsistent article pages, or missing metadata can reduce the reliability of article identification. The findings therefore indicate that metadata quality should be understood as part of the technical infrastructure of scholarly communication rather than merely an editorial formatting issue.

3.3.3 Changes in Journal Websites and Digital Archives

Changes to journal websites and digital archives emerged as an additional source of indexing instability. Website migration, changes in hosting providers, restructuring of journal platforms, modification of article URLs, removal of older issues, and relocation of PDF files can interrupt the relationship between previously indexed records and their current online locations. Google Scholar's guidance emphasizes that crawlers must be able to discover article URLs and periodically refresh their content (Islam, 2025). Each paper should have its own unique URL, while article pages should remain accessible through a navigable website structure.

The implications are particularly significant for journals that migrate from one publishing platform to another without preserving old URLs or implementing appropriate redirects. An article that was previously accessible at one location may become unavailable when its URL changes, even though the article itself has not been withdrawn. This distinction is important because apparent de-indexing may therefore represent a failure of digital continuity rather than an editorial decision. Evidence from the literature supports this interpretation. Stainforth (2022) documented substantial fluctuations in Google Scholar coverage and observed that missing documents could later reappear following subsequent index updates. Google also acknowledges that changes to already indexed content can take considerably longer to be reflected than newly published material.

The findings consequently highlight the importance of stable digital archives. Journal publishers should preserve complete back issues, maintain persistent article URLs, retain searchable PDF files, and carefully manage website migrations. A journal that repeatedly changes its technical environment without preserving article-level continuity increases the probability of broken links,

inaccessible historical content, incomplete archive coverage, and inconsistent indexing (Maulidar, 2026). Sustainable indexing therefore requires digital preservation to be integrated into journal publishing policies rather than treated solely as an information-technology concern.

3.4 Identification, Diagnosis, and Verification of De-Indexing

3.4.1 Indicators of Possible De-Indexing

The review identified several observable indicators that may raise suspicion of Google Scholar de-indexing. These include the disappearance of previously searchable articles, a sudden decline in the number of articles associated with a journal, missing citation records, incomplete coverage of particular issues or years, and substantial reductions in the discoverability of articles that remain available on the journal website (Popov, 2018). A particularly important indicator is the disappearance of multiple previously indexed articles within a relatively short period, because this pattern may suggest a broader indexing or technical problem rather than an isolated search-result fluctuation.

Nevertheless, the evidence indicates that these indicators should be interpreted cautiously. Google Scholar itself acknowledges that gaps in coverage can result from technical issues affecting the automatic processing of websites. Empirical research also demonstrates that Google Scholar coverage can fluctuate considerably over time. Teixeira da Silva (2023) documented a major coverage reduction affecting scholarly documents from several publishers, followed by subsequent recovery. Therefore, disappearance from an individual search result should be regarded as evidence of possible visibility loss rather than definitive proof of formal de-indexing.

3.4.2 Distinguishing De-Indexing from Temporary or Technical Problems

A central finding of the review is that disappearance from Google Scholar should not automatically be interpreted as permanent de-indexing. Several alternative explanations can produce the same observable outcome. These include temporary website downtime, crawler failures, changes to metadata, inaccessible PDFs, URL restructuring, duplicate-record consolidation, delays in processing new or modified content, and fluctuations in Google's indexing systems (Tavani, 2012). Google Scholar indicates that it normally adds new papers several times a week but that larger websites may take longer to crawl, while corrections to existing records can take considerably longer.

The literature provides particularly strong evidence for distinguishing technical disruption from deliberate exclusion. The Astronomy and Astrophysics case studied by Askar (2025) found that documents disappeared despite continuing to be available on the Web, and that many later returned following an index-wide update. This finding demonstrates that apparent de-indexing can arise from temporary or systemic indexing failures. Similarly, research on Google Scholar data has documented duplicates, missing URLs, inaccessible documents, incorrect author information, title errors, and publication-data errors, all of which can alter the apparent visibility of scholarly works without necessarily representing formal removal.

Accordingly, the findings support a distinction between confirmed de-indexing, temporary indexing disruption, and record-level indexing problems. This distinction is methodologically important because treating every missing search result as de-indexing may lead publishers and researchers to undertake unnecessary corrective actions or make unsupported claims about Google Scholar's decisions (Bhattacharjee, 2023).

3.4.3 Procedures for Verification and Diagnosis

The review indicates that suspected de-indexing should be investigated through a systematic verification process rather than through a single Google Scholar search. The first step is to search for the exact article title, followed by searches using combinations of the author's name, journal title, publication year, DOI or other publication identifier. Searching the journal title together with distinctive article terms can also establish whether other articles from the same issue remain visible (Yoon, 2014). These approaches help determine whether the problem concerns an individual article, a group of articles, a particular issue, or the journal as a whole.

Verification should then be extended to the journal's own website. The investigator should determine whether the article page and PDF remain publicly accessible, whether the URL has changed, whether the PDF contains searchable text, and whether the article's bibliographic metadata remains correctly represented. Google Scholar specifically recommends checking article pages and their HTML metadata when troubleshooting indexing problems (Pandita, 2022). Comparing current visibility with previously recorded search results, citation records, or archived journal records can provide additional evidence concerning when the disappearance occurred.

The findings therefore support a diagnostic sequence based on search verification, website verification, metadata verification, archive comparison, and temporal monitoring. This approach is preferable to relying on a single search observation. It also

recognizes that Google Scholar coverage is dynamic (Cortegiani, 2020). The documented case of substantial coverage loss followed by later recovery demonstrates why longitudinal monitoring is particularly valuable when diagnosing suspected de-indexing.

3.5 Consequences and Implications of Google Scholar De-Indexing

3.5.1 Effects on Authors, Researchers, and Citation Visibility

The review indicates that reduced Google Scholar visibility can have important consequences for authors because discoverability is closely connected to the opportunity for research to be read, cited, and incorporated into subsequent scholarship. When an article becomes difficult to retrieve, researchers searching by title, author, or subject may fail to encounter it even when the article remains available on the publisher's website (Khatri, 2023). This can reduce opportunities for readership and citation accumulation and may create incomplete representations of an author's scholarly output.

The implications extend to researcher profiles and bibliometric indicators. Google Scholar aggregates different versions of scholarly works and uses bibliographic information to connect citations with publications. When records disappear, are duplicated, or contain inaccurate metadata, citation counts and publication lists may become unstable. Google Scholar itself notes that citation counts can decline when papers or citing papers disappear from the Web, become unavailable to crawlers, or are reformatted in ways that make bibliographic information and references difficult for automated systems to identify (Mira, 2022). Thus, de-indexing or indexing disruption can affect not only discoverability but also the reliability of indicators used by researchers to demonstrate scholarly visibility.

3.5.2 Effects on Journals, Publishers, and Scholarly Communication

At the journal level, persistent indexing problems can affect readership, citation performance, author confidence, and perceptions of publishing quality. A journal whose articles repeatedly disappear from Google Scholar may become less discoverable to researchers even when its editorial operations remain active (Frosio, 2020). Reduced discoverability can consequently weaken the circulation of published research and make the journal less visible within the broader scholarly communication ecosystem.

The findings also suggest that indexing problems can undermine confidence in a publisher's digital infrastructure. Authors generally expect published articles to remain accessible and discoverable over time. When journal migrations result in broken links, missing back issues, or inconsistent article records, authors may perceive the publishing environment as unstable. This is particularly important because Google Scholar is widely used for scholarly discovery, although it is not a definitive measure of research quality (Savegnago de Mira, 2022). The literature demonstrates that Google Scholar itself contains substantial data-quality problems and coverage fluctuations, meaning that journals should not equate indexing status with scholarly legitimacy. Nevertheless, persistent visibility problems remain significant because they can reduce the practical reach of otherwise legitimate scholarly work.

3.5.3 Recovery, Prevention, and Sustainable Indexing Practices

The findings indicate that effective management of Google Scholar indexing requires a combination of technical maintenance, metadata quality, editorial consistency, digital preservation, and ethical publishing practices. At the technical level, journals should maintain publicly accessible article pages, searchable PDFs, stable URLs, crawlable archive structures, and properly configured bibliographic metadata. Google Scholar specifically recommends structured metadata, unique article URLs, accessible abstracts or full texts, and website structures that allow crawlers to discover individual articles (Singson, 2025). Publishers should therefore test article pages regularly rather than assuming that successful publication automatically guarantees continued indexing.

Prevention should also include systematic monitoring. Journals can periodically search representative articles from each issue and publication year using exact titles, authors, journal names, and publication identifiers. A record of previously visible articles can provide a baseline against which future changes are assessed. Where visibility declines, publishers should first check whether articles remain accessible, whether URLs or hosting arrangements have changed, and whether metadata and document structures remain intact (Ananny, 2016). This approach is consistent with evidence showing that Google Scholar coverage can experience substantial temporary fluctuations and that recovery may occur after later indexing updates.

Overall, the review demonstrates that sustainable Google Scholar indexing cannot be reduced to a single technical configuration. It depends on the interaction of editorial quality, technical reliability, digital preservation, metadata accuracy, and publication ethics. Technical accessibility enables discovery; accurate metadata enables identification and citation matching; stable archives preserve continuity; and ethical editorial practices support the credibility of the scholarly record (Sookman, 2018). Consequently, publishers seeking to minimize indexing disruptions should adopt a continuous quality-assurance approach that combines

website monitoring, metadata validation, archive preservation, article-level verification, and transparent editorial management. Such an integrated approach is more sustainable than treating Google Scholar visibility as a one-time registration or indexing event.

4. Conclusion

This review concludes that Google Scholar de-indexing is a multifaceted phenomenon arising from the interaction of editorial, technical, and ethical factors rather than from a single cause. The findings indicate that inadequate editorial practices, inconsistent publication information, questionable or non-transparent publishing standards, website accessibility problems, crawler incompatibility, unstable article URLs, poorly structured metadata, and changes in indexing policies can all contribute to the disappearance or reduced visibility of scholarly content in Google Scholar. Ethical concerns, particularly plagiarism, duplicate publication, manipulated citation practices, fabricated content, and other forms of scholarly misconduct, may further undermine the credibility and continued discoverability of journals and articles. The review therefore emphasizes that maintaining Google Scholar visibility requires publishers and journal editors to adopt robust editorial quality-control procedures, ensure technically accessible and well-structured journal websites, provide accurate and consistent metadata, and uphold internationally recognized principles of research and publication ethics. De-indexing should consequently be understood not merely as a technical indexing problem but as an important indicator of the broader quality, reliability, transparency, and sustainability of scholarly publishing practices. Proactive monitoring, regular technical audits, ethical compliance, and prompt correction of identified problems can help journals reduce the risk of de-indexing and strengthen their long-term visibility, credibility, and accessibility within the scholarly communication ecosystem.

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] Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, *41*(1), 93-117.

[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] Askar, N. (2025). The Right to Be Forgotten in the Age of Artificial Intelligence. In *European Yearbook on Human Rights 2025* (pp. 451-480). Brill Nijhoff.

[4] Bhattacharjee, M., Kaposy, C., Grossman, M. R., & Marshall, Z. (2023). Patient photographs on Google Images: a commentary on informed consent, copyright, and privacy laws. *Law, Innovation and Technology*, *15*(2), 536-557.

[5] Brock, G. (2016). *The right to be forgotten: privacy and the media in the digital age*. Bloomsbury Publishing.

[6] 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.

[7] Frosio, G. (Ed.). (2020). *Oxford handbook of online intermediary liability*. Oxford University Press.

[8] Gregory, S. (2018). Mal-uses of AI-generated Synthetic Media and Deepfakes, Pragmatic Solutions Discovery Convening, June 2018: Summary of Discussions and Next Step Recommendations. *Mal-uses of AI-generated Synthetic Media and Deepfakes, Pragmatic Solutions Discovery Convening, June 2018: Summary of Discussions and Next Step Recommendations*.

[9] Islam, M. A. (2026). Constraint-Responsive Human Behavior for Organizations: A Conceptual Theory. *South Asian Journal of Social Studies and Economics*, *23*(6), 44–52. https://doi.org/10.9734/sajsse/2026/v23i61329

[10] Islam, M. A., & Aktar, L. (2025). Perceived Ease of Use, Security, and Trust as Predictors of Online Purchase Intention: A Technology Acceptance Model Extension. *European Economics Letters*, *15*(3).

[11] Islam, M. A., Aktar, N., Barua, P., Sweety, M. A., Aktar, L., & Islam, M. B. (2025). Perceived Competitiveness in Malaysian Higher Education: Role of International Student Recruitment Strategies. *Asian Journal of Education and Social Studies*, *51*(9), 997-1011.

[12] Islam, M. A., Jantan, A. H. B., Islam, M. A., Abdullah, A. B. M., & Rahman, M. S. (2026). Unlocking the Dynamics of Employee Retention: Examining the Interplay of Job Security, Promotion and Work Engagement in a Developing Economy. *FIIB Business Review*, 23197145261431894. DOI: 10.1177/23197145261431894.

[13] Khatri, S., & Fatma, T. (2023). Second Chances and Digital Erasure: Do Former Convicts Have the Right to Be" Forgotten" in India?. *CALI*, *8*, vii.

[14] Lyseiuk, A. (2025). The Right to Be Forgotten vs. the Right to History: A Communicational Conflict between Personal Data Protection and the Preservation of Public Digital Archives. *J. Int'l Legal Commc'n*, *18*, 77.

[15] Marinett, M. T. (2024). *Corporate Responsibility and Accountability in Internet Content Governance* (Doctoral dissertation, University of Toronto (Canada)).

[16] Maulidar, I. (2026). Anatomy of Unpublishing: Negotiation of Un (Facts) in Newsroom Practice. *Journalism Practice*, 1-19.

[17] 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.
[18] O'Hara, K., Shadbolt, N., & Hall, W. (2016). A pragmatic approach to the right to be forgotten.
[19] Pandita, R., Koul, M., & Singh, S. (2022). Indexing and de-indexing of journals in UGC CARE list: A critical commentary. *University News*, *60*(19), 7-12.
[20] Pasquale, F. (2020). Internet nondiscrimination principles revisited. *Brooklyn Law School, Legal Studies Paper*, (655).
[21] Popov, G., Antwi, S., Onyederi, J., James, I., Brotherton, D., & Studdard, B. (2018). Texas Undergraduate Law Review Journal at The University of Texas at Austin, Volume VII, Issue I.
[22] 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.
[23] Singson, M., Manavalan, L., & Thiyagarajan, S. (2025). Analysing the impact of academic policies and interventions on research productivity in Indian library and information science. *Performance Measurement and Metrics*, *26*(1), 17-33.
[24] Slane, A. (2018). Search engines and the right to be forgotten: Squaring the remedy with Canadian values on personal information flow. *Osgoode Hall LJ*, *55*, 349.
[25] Sookman, B. (2018). US Court Thumbs Its Nose at Supreme Court of Canada: Google v. Equustek. *Intellectual Property Journal*, *30*(3), 375-405.
[26] Stainforth, E. (2022). Collective memory or the right to be forgotten? Cultures of digital memory and forgetting in the European Union. *Memory Studies*, *15*(2), 257-270.
[27] Tavani, H., & Zimmer, M. (2012). Search engines and ethics.
[28] 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.
[29] 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.
[30] Yoon, Y. J. (2014). Legal Ramifications of" Black Hat" Search Engine Optimization. *BUJ Sci. & Tech. L.*, *20*, 159.
[31] Zhao, Z. (2022). The application of the right to be forgotten in the machine learning context: From the perspective of european laws. *Cath. UJL & Tech*, *31*, 73.