



Algorithmic Bias and Information Trust in Academic Library Discovery Systems: An Analytical Study of User Search Behaviour and Digital Information Literacy Using Online Primary Empirical Evidence

¹Sunil Kumar, ²Dr Rajeev Manhas, ³Mamta, ⁴Ruchi Dhiman and ⁵Chand

¹Chief Librarian Punjab Institute of Medical Sciences Medical College and Hospital Jalandhar Punjab
Email ID:- librariansunilkumar1@gmail.com Orchid ID:- <https://orcid.org/0009-0005-3724-3481>

²Chief Librarian, Baba Farid University of Health Sciences, Faridkot Punjab
Email ID : dr.rajeevmanhas@gmail.com Orcid ID- <https://orcid.org/0000-0002-3996-2421>

³PhD Research Scholar , Department of Library and Information Science, HRIT University Ghaziabad, UP,
India
Mail. ID:- mamta116kumari@gmail.com

⁴PhD Research Scholar , Department of Library and Information Science, Sparsh Himaliyaya University,
Dehradun, Utrakhnad Emai ID- ruchi.libsc@gmail.com Orcid ID:- <https://orcid.org/0009-0005-4922-7788>

School Librarian, Darshan Academy Senior Secondary School, Jalandhar, Punjab
Mail. ID:- chandpul8@gmail.com

How to Cite this Article:

Kumar, S., Manhas, D. R., Mamta, , Dhiman, R. & Chand, (2026). Algorithmic Bias and Information Trust in Academic Library Discovery Systems: An Analytical Study of User Search Behaviour and Digital Information Literacy Using Online Primary Empirical Evidence. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(10), 1-9.
<https://doi.org/10.55041/ijcope.v2i10.093>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.093>

Abstract

Background: Academic library discovery systems increasingly mediate access to scholarly information through ranking, metadata matching, filtering, database coverage, availability signals, and recommender functions. These algorithmic arrangements can shape what users notice, cite, and trust. **Objective:** This paper examines how algorithmic bias, user search behaviour, information trust, and digital information literacy intersect in academic discovery environments. **Methodology:** Since no local institutional dataset was provided, the paper uses a secondary analysis of online primary empirical evidence from three verifiable sources: Project Information Literacy's student survey of 2,318 respondents across six U.S. campuses; a published Summon query/click-log analysis based on anonymized searches by millions of users at hundreds of libraries; and an open representative German online survey on search-engine literacy and trust involving 2,012 respondents. **Results:** The data show high dependence on scholarly databases and OPACs but limited engagement with librarians. In the Project Information Literacy survey, 84% of students used scholarly research databases and 78% used OPACs for course-related research, while only 20% consulted librarians about assignments. Summon logs show that users generally began with simple search, with fewer than 6% of queries using fielded search and only slightly more than 2% using advanced operators,

although more than 40% used at least one filter. Search-engine literacy data show that 79% knew paid ads exist, but only 29.2% knew how ads differ from organic results, and only 8.9% knew the term SEO. **Implications:** Libraries need



discovery-literacy instruction that connects searching, ranking, metadata quality, database coverage, and calibrated trust.

Keywords: algorithmic bias; academic libraries; discovery systems; information trust; search behaviour; digital information literacy; information retrieval

1. Introduction

Academic library discovery systems have transformed scholarly searching from a fragmented process of consulting separate catalogues, indexes, repositories, and licensed databases into an integrated search experience. Systems such as Primo, EBSCO Discovery Service, Summon, WorldCat Discovery, Koha OPAC integrations, institutional repositories, Google Scholar, and subject databases now operate as central gateways to academic information. Their value lies in convenience and scale, but their operation is not neutral: ranking, metadata mapping, record deduplication, database coverage, availability labels, recommender links, and interface defaults shape the order and visibility of scholarly material (Asher et al., 2013; Bossaller & Sandy, 2017; Dahlen & Hanson, 2017).

The algorithmic character of discovery systems creates a new problem for Library and Information Science. Users often experience the ranked result list as a natural or authoritative order, whereas it is a computational arrangement derived from keyword matching, metadata completeness, full-text indexing, local holdings, citations, usage signals, vendor configuration, and interface design. Friedman and Nissenbaum (1996) argued that bias in computer systems may be pre-existing, technical, or emergent. This framework is directly applicable to library discovery because inequities in publishing, language dominance, incomplete metadata, uneven coverage of regional scholarship, and user click behaviour can all become signals that influence search visibility.

Information trust is therefore not produced only by the quality of individual sources. In discovery environments, trust is also produced by institutional branding, ranking position, database labels, full-text links, metadata richness, peer-review facets, and familiar interface cues. Rieh (2002) showed that users judge web information through both information-quality judgments and cognitive authority. In academic libraries, the library's trusted identity can mask the vendor, collection, and metadata decisions behind the result list (Reidsma, 2019).

Digital information literacy is a key mediator in this process. The ACRL Framework emphasizes that searching is strategic exploration and that authority is constructed and contextual (Association of College and Research Libraries, 2016). Metaliteracy extends this position by stressing critical participation in networked information environments (Mackey & Jacobson, 2011). In the context of discovery systems, digital information literacy means more than knowing how to type a keyword: it includes knowing how ranking works, when to compare databases, how to interpret metadata, how to evaluate source authority, and when top-ranked results require verification.

The present study responds to a practical and conceptual gap. Prior studies have examined discovery-layer performance, student search behaviour, search-engine trust, and algorithmic bias, but these areas are often treated separately. This paper integrates them through a data-based analytical approach. Because no local field data were supplied, it uses real online primary empirical sources rather than fabricated survey or interview results. The paper therefore provides a publishable methodology and descriptive evidence base while remaining transparent about the limits of secondary analysis.

1.1 Research Gap

Existing literature has established that discovery systems improve access but vary in retrieval effectiveness and disciplinary coverage (Asher et al., 2013; Hanneke & O'Brien, 2016; Pulikowski & Matysek, 2021). Studies of search behaviour also show that students rely heavily on convenient and familiar search tools (Head & Eisenberg, 2009; Holman, 2011). However, fewer studies empirically connect algorithmic ranking, metadata quality, database coverage, and information trust with users' digital information literacy in a single analytical framework. The research gap is especially clear for studies that compare self-reported trust with observed search behaviour and actual discovery-log evidence.

1.2 Research Objectives

- To examine how ranking, metadata quality, filtering, database coverage, and recommender functions influence users' trust in academic discovery environments.



- To analyze real online primary evidence on student use of scholarly databases, OPACs, librarian support, and one-search features.
- To examine large-scale discovery-log evidence on query formulation, fielded search, advanced operators, filter use, and click concentration.
- To assess how search-engine and algorithmic literacy indicators relate to information trust and awareness of ranking influence.
- To develop practical implications for library instruction, metadata governance, discovery-system evaluation, and user education.

1.3 Research Questions

- What do available online primary data show about students' use of scholarly databases, OPACs, and librarian support for academic research?
- What patterns of search behaviour are visible in discovery-system query logs?
- What do search-engine literacy data suggest about users' awareness of ranking influence, paid placement, and organic results?
- How do credibility, depth, one-search convenience, metadata, and interface simplicity shape information trust?
- What forms of digital information literacy instruction are required to reduce blind trust in top-ranked discovery results?

1.4 Analytical Propositions

- P1: Users' trust in discovery systems is shaped by institutional authority, convenience, and ranking position, not only by independent source evaluation.
- P2: Heavy use of scholarly databases does not automatically imply strong awareness of algorithmic ranking or discovery-system bias.
- P3: Search behaviour is often simple and convenience-oriented, with limited use of advanced search syntax, even in academic search environments.
- P4: Digital information literacy should include algorithmic awareness, metadata interpretation, database comparison, and critical evaluation of top-ranked results.

2. Review of Literature

2.1 Algorithmic Bias in Information Systems

Algorithmic bias is best understood as a socio-technical phenomenon rather than as a mere software error. Friedman and Nissenbaum (1996) identify pre-existing bias, technical bias, and emergent bias. Pre-existing bias enters discovery systems through broader inequalities in publishing, language, indexing, and scholarly visibility. Technical bias enters through metadata mapping, ranking formulas, deduplication rules, availability weighting, and recommender logic. Emergent bias appears when user behaviour, such as repeated clicking on top-ranked results, feeds back into relevance models or popularity signals.

The politics of search has long been recognized in information ethics. Introna and Nissenbaum (2000) argued that search engines structure public visibility and therefore perform political work. Noble (2018) demonstrated that ranking systems can reproduce social inequalities and stereotypes. Although academic library discovery systems differ from commercial web search engines, they share the same basic power: they allocate attention. What is highly ranked is more likely to be opened, read, cited, and treated as authoritative.

2.2 Bias in Search Engines, Ranking, and Recommender Systems

Ranking bias occurs when users treat higher-ranked results as more relevant or more credible than lower-ranked results. Personalization and filtering can also limit exposure to diverse sources (Bozdag, 2013; Hannak et al., 2013). In academic search, popularity bias may privilege frequently cited, frequently used, or commercially indexed literature, while newer, regional, non-English, interdisciplinary, or poorly indexed research remains less visible.

The representative survey evidence on search-engine trust is relevant because library discovery interfaces increasingly resemble web search interfaces. Schultheiss and Lewandowski (2023) found that users with less search-engine



knowledge may be more likely to trust and use Google, revealing a mismatch between confidence and evaluative ability. This pattern is analytically relevant for academic discovery systems, where library branding may increase trust even when users do not understand ranking and coverage limits.

2.3 Academic Library Discovery Layers

Discovery layers were designed to reduce fragmentation in academic searching. Asher et al. (2013) showed that discovery tools, Google Scholar, and conventional library resources differ in effectiveness, meaning no single tool can fully solve the scholarly discovery problem. Hanneke and O'Brien (2016) similarly showed that web-scale discovery performance depends on disciplinary context and configuration. Dahlen and Hanson (2017) found that students' preferred search tools may conflict with subject-specific authority, while Pearce (2019) showed that discovery tools may affect disciplines differently.

Bossaller and Sandy (2017) called for more systematic evaluation of discovery layers. This is crucial because discovery systems do not merely retrieve information; they make collection, metadata, ranking, and interface decisions visible or invisible to users. The less visible these decisions are, the harder it becomes for users to calibrate trust.

2.4 User Search Behaviour in Digital Libraries

Student search behaviour is frequently shaped by familiarity, time pressure, assignment expectations, and prior training. Project Information Literacy found that students often rely on familiar sources and routines for both course-related and everyday-life research (Head & Eisenberg, 2009). Holman (2011) also emphasized that millennial students bring search-engine mental models into academic database use, which can influence how they interpret search boxes, result lists, and filters.

Large-scale search logs add behavioural evidence beyond self-reports. Diamond et al. (2013) argue that query logs capture what users actually do rather than what they say they do. Their Summon log analysis shows that most users begin with simple searches, only a small share use fielded or advanced syntax, and filters are more common than advanced operators. This evidence helps explain why discovery interfaces must support both simplicity and critical scaffolding.

2.5 Information Trust and Credibility Assessment

Information trust involves confidence that retrieved information is relevant, credible, authoritative, and suitable for academic use. Rieh (2002) distinguishes between judgments of information quality and cognitive authority. In discovery systems, authority is inferred through author names, journal titles, database labels, peer-review indicators, abstracts, citation links, availability labels, and rank position. When these cues are incomplete or opaque, trust becomes vulnerable to miscalibration.

Reidsma (2019) argues that bias in library discovery can be masked by user trust in the library. This is a critical point: library-branded systems may be trusted more than their underlying vendor algorithms and metadata quality justify. A mature trust model therefore requires users to trust the library as an institution while still critically evaluating the system's ranking, coverage, and metadata limitations.

2.6 Digital Information Literacy

Digital information literacy includes technical, cognitive, and evaluative capabilities (Bawden, 2008). The ACRL Framework emphasizes authority, information creation, scholarship as conversation, strategic exploration, and information value (Association of College and Research Libraries, 2016). UNESCO's media and information literacy framework similarly stresses critical judgment in digitally mediated information environments (UNESCO, 2021).

In discovery-system use, digital information literacy must include algorithmic literacy. Users need to know that relevance ranking is calculated, database coverage is partial, metadata quality varies, and one-search convenience may hide disciplinary or linguistic gaps. Librarians therefore need to teach not only which buttons to click but also how scholarly visibility is produced.

2.7 Metadata, Coverage, Language, and Popularity Bias

Metadata quality affects whether scholarly material can be found and trusted. Incomplete titles, inconsistent author names, missing abstracts, weak subject headings, duplicate records, broken links, and poor language representation can reduce relevance and visibility. Bowker and Star (1999) showed that classification systems embody institutional



assumptions, while Olson (2002) demonstrated that subject representation can limit whose knowledge is nameable and retrievable.

Coverage bias occurs when certain publishers, countries, languages, disciplines, or formats are overrepresented in discovery indexes. Popularity bias occurs when usage, citation, or click signals favour already visible materials. Language bias is particularly important in multilingual and Global South contexts, where English-language and commercially indexed materials may dominate result lists. These biases interact with ranking and trust because users may confuse the boundaries of a platform with the boundaries of knowledge.

3. Methodology

3.1 Research Design

The revised study uses a secondary analysis of online primary empirical data. This design is necessary because no new survey responses, interview transcripts, observation sheets, or local search-log file were provided by the researcher. The paper therefore uses only genuine online empirical sources that report or provide primary data relevant to academic search behaviour, discovery-system use, search-engine literacy, and information trust. The analysis is descriptive and interpretive rather than inferential, because the included sources were collected in different populations and cannot be merged into a single respondent-level dataset without violating methodological assumptions.

3.2 Online Primary Data Sources

Table 1. Online Primary Empirical Data Sources Used in the Revised Study

Source	Population/context	Sample/coverage	Data type	Variables used
Project Information Literacy survey (Head & Eisenberg, 2009)	College students from six U.S. campuses	2,318 respondents; item Ns vary by question	Online survey and follow-up evidence	Use of scholarly databases, OPAC, librarians, and reasons for database trust
Summon Relevance Metrics Framework log analysis (Diamond et al., 2013)	Users of the Summon discovery service across many libraries	Millions of users at hundreds of libraries; aggregate log results reported	Anonymized query and click-through logs	Simple search, fielded search, advanced operators, filters, click concentration, search sessions
SEO Effect / F1000Research and OSF survey dataset (Schultheiss & Lewandowski, 2022; Lewandowski & Schultheiss, 2022)	Representative German online population aged 16-69	N = 2,012; large screen n = 999; small screen n = 1,013	Open survey dataset and published aggregate results	Search-engine literacy, trust, awareness of ads, SEO awareness, ranking influence

Note. The study uses these sources as online primary empirical evidence. The PIL and Summon sources are directly connected to academic libraries and discovery systems. The SEO Effect survey is used cautiously as comparable evidence on algorithmic search literacy and trust.

The Project Information Literacy source is directly relevant to academic library behaviour because it reports students use of scholarly databases, OPACs, librarian support, and reasons for using online scholarly research databases. The Summon log analysis is directly relevant because it reports anonymized behaviour in a large-scale library discovery system. The SEO Effect survey is not an academic-library dataset, but it is relevant because it measures user awareness of search-engine ranking, ads, SEO, trust, and personalization--constructs closely related to algorithmic awareness in library discovery environments.



3.3 Variables and Operational Definitions

Table 2. Variables and Operational Definitions

Variable group	Operational definition	Empirical source
Search behaviour	Use of scholarly databases, OPACs, filters, fielded search, advanced operators, librarian consultation, search-session behaviour	PIL survey; Summon logs
Digital information literacy	Ability to select databases, compare sources, use filters/fields, understand ranking influence, distinguish paid/organic results, and evaluate credibility	PIL survey; F1000/OSF survey; Summon logs
Algorithmic awareness	Recognition that ranking can be influenced by algorithms, SEO, ads, metadata, filters, coverage, and interface defaults	F1000/OSF survey; Summon logs
Information trust	Use of credibility cues such as database authority, peer review, in-depth content, institutional expectation, correct/trustworthy search results, and rank position	PIL survey; F1000/OSF survey
Discovery-system bias	Potential bias from ranking position, metadata quality, database coverage, language coverage, popularity, personalization, and recommender features	Literature synthesis; Summon logs; F1000/OSF survey

Note. The variables are operationalized for secondary analysis. A future local field survey can use the same constructs for institution-specific hypothesis testing.

3.4 Inclusion and Exclusion Criteria

- Included sources had to contain real empirical data related to academic search behaviour, discovery-system logs, search-engine literacy, or information trust.
- Included values had to be reported by an identifiable scholarly, institutional, or open-data source with accessible documentation.
- Excluded sources included opinion articles without empirical data, commercial promotional materials, unverifiable datasets, and any dataset lacking clear methodological description.
- Only published or documented numerical values were extracted. Inferential statistics were not created from aggregate tables.

3.5 Ethical Considerations

This revision uses only publicly available, open, or published aggregate empirical data. The Summon log source reports anonymized query and click-through analysis and does not identify end users. The F1000/OSF dataset is documented as available under a Creative Commons Attribution 4.0 license. No personal information from individual respondents has been accessed or reported in this paper. For any future local study, institutional ethics approval, informed consent, anonymization, and responsible search-log governance will be required.



4. Data Collection Procedure

Data for this revised paper were collected through documentary extraction from online primary empirical sources. The extraction process followed five steps. First, sources were screened for empirical relevance to library discovery, search behaviour, algorithmic ranking, digital information literacy, and information trust. Second, sample size, population, study design, and data type were recorded. Third, numerical values were extracted only from published tables, methodology sections, abstracts, or documented findings. Fourth, each extracted value was mapped to one of the study constructs: search behaviour, information trust, algorithmic awareness, digital information literacy, or system transparency. Fifth, the values were inserted into the Results section with source notes.

The extraction method is conservative. The paper uses Project Information Literacy figures on the proportion of students using scholarly databases, OPACs, librarians, training sessions, and database trust cues. It uses Summon log statistics on fielded search, advanced operator use, filter use, and click concentration. It uses the SEO Effect survey's reported values on awareness of paid ads, knowledge of how ads differ from organic results, knowledge of SEO, and perceived impact of SEO on rankings. Values not provided by the sources are not inferred.

Table 3. Data Extraction Procedure

Step	Action	Output
Source identification	Locate empirical sources connected to discovery systems, academic search, search logs, or search-engine literacy	Selected three data-bearing online sources
Methodological screening	Verify sample, population, data type, and availability	Excluded unsupported claims and non-empirical sources
Numerical extraction	Extract only reported values and item Ns	Created data tables without invented values
Construct mapping	Assign values to search behaviour, trust, literacy, and awareness constructs	Produced integrated results matrix
Integrity check	Avoid p-values, alpha values, and regression coefficients where raw comparable data are unavailable	Maintained statistical transparency

Note. This table documents how online primary empirical values were moved into the revised manuscript.

5. Data Analysis and Statistical Techniques

The analysis uses descriptive statistics and cross-source synthesis. Frequencies and percentages are reported exactly as they appear in the selected sources. Where sources report item-specific valid sample sizes, those Ns are retained rather than replaced by a single artificial N. Because the sources use different populations and instruments, the paper does not merge the data into one dataset and does not calculate new p-values, Cronbach's alpha, correlations, ANOVA, or regression coefficients. Doing so from published aggregate values would be methodologically invalid and would risk fabricating findings.

The statistical interpretation proceeds in three layers: first, student academic-library behaviour is described using Project Information Literacy values; second, discovery-system behaviour is described using Summon query-log indicators; third, algorithmic and search-engine literacy is described using SEO Effect survey findings. These descriptive results are then interpreted through the literature on algorithmic bias, information trust, and digital information literacy.



Table 4. Statistical Technique Status

Technique	Status in this paper	Reason/source	Integrity decision
Frequency and percentage	Used	PIL survey tables; SEO Effect reported percentages; Summon reported percentages	Report actual values only
Mean and standard deviation	Not recalculated	Raw item-level data not processed in this revision	Avoid fabricated central tendency values
Cronbach alpha	Not reported	Scale-level raw responses not analyzed here	Can be computed only after raw survey dataset is processed
Correlation/regression	Not reported	Sources are not one merged respondent-level dataset	Do not infer relationships from aggregate tables
Thematic interpretation	Used	Literature plus empirical patterns	Interpret descriptively, not causally

Note. Inferential statistics are intentionally not invented. Descriptive statistics are filled using real online primary empirical evidence.

6. Statistics and Results

6.1 Student Use of Academic Library Resources

The Project Information Literacy survey provides directly relevant primary evidence on students' use of library resources for course-related research. The strongest pattern is that students used online scholarly databases and OPACs more than librarian-mediated services. This indicates a search environment in which discovery tools and databases carry major responsibility for shaping trust and relevance judgments.

Table 5. Usage of Library Resources and Services for Course-Related Research

Resource/service used	Total used (n)	Valid N	Percentage
Use scholarly research databases (EBSCO, JSTOR, ProQuest)	1,866	2,216	84%
Use online public access catalog (OPAC)	1,729	2,222	78%
Use library study areas	1,596	2,222	72%
Look on library shelves for materials	1,197	2,193	55%
Use library cafe	1,045	2,171	48%
Ask a librarian about workings of the library system	527	2,212	24%
Consult a librarian about a course-related assignment	446	2,227	20%



Resource/service used	Total used (n)	Valid N	Percentage
Attend non-credit library training session	274	2,193	12%
Use Ask-a-Librarian chat, email, or IM	263	2,189	12%

Note. Source: Head and Eisenberg (2009), Figure 10. Values are item-specific and therefore valid Ns vary by row.

The data show that database and catalogue access dominated students' library-related search behaviour, while direct librarian consultation and training were much lower. This pattern matters for algorithmic trust because many users encounter the scholarly record through interfaces and rankings before they encounter expert mediation by librarians.

6.2 Database Trust and Relevance Cues

The same Project Information Literacy survey reports reasons students used scholarly research databases. These values show that students associated scholarly databases with credibility, depth, instructor expectations, and convenience. This is direct evidence that database use is tied to information trust, not merely access.

Table 6. Reasons for Using Scholarly Research Databases

Reason for database use	Total used (n)	Valid N	Percentage
Have more credible content than what might be found on the Internet	1,735	2,220	78%
Have in-depth, detailed information	1,676	2,218	76%
Have the kind of resources my instructors expect to see	1,642	2,220	74%
Have an interface that makes it easy to find sources	1,436	2,217	65%
Have allowed me to succeed in the past	1,368	2,222	62%
Have a one-search feature	1,342	2,225	60%
Require no visit to the library building	1,235	2,222	56%
Have peer-reviewed sources	1,124	2,218	51%
Allow quick just-in-time article discovery	941	2,218	43%

Note. Source: Head and Eisenberg (2009), Figure 13. These reasons function as empirical indicators of information trust and perceived relevance.

The high value attached to credible content (78%) and in-depth information (76%) suggests that students treat library databases as trusted academic environments. However, the 60% value for the one-search feature also shows that convenience and interface design are part of trust formation. If a one-search interface hides database coverage gaps or rank logic, users may trust convenience without recognizing mediation.



6.3 Course-Related Research Tasks

Table 7. Course-Related Research Assignments Reported by Students

Assignment type	Total (n)	Valid N	Percentage
Argument paper about an issue	1,518	2,266	67%
Interpretative reading of a text with research from other sources	1,203	2,266	53%
Historical analysis of an event	878	2,266	39%
Literature review	863	2,266	38%
Case study analysis	748	2,266	33%
No experience with course-related research papers at that institution	204	2,266	9%

Note. Source: Head and Eisenberg (2009), Figure 1. The table shows the academic tasks for which discovery systems and databases may be used.

6.4 Discovery-System Query Behaviour from Summon Logs

The Summon Relevance Metrics Framework provides large-scale behavioural evidence from anonymized query and click-through logs. Unlike survey data, log evidence captures search actions. The reported values show that users often begin with simple queries and use advanced functionality selectively.

Table 8. Summon Query-Log Indicators of Discovery-System Search Behaviour

Indicator	Reported value	Analytical meaning
Clicked-result concentration	Most clicked documents appear in the top 10 results	Supports the importance of rank position and first-page visibility
Fielded search in queries	Less than 6% of queries specify one or more search fields	Advanced bibliographic structuring is uncommon
Sessions beginning with fielded search	Less than 4% of sessions start with search fields	Most users begin with simple search
Advanced operators	A little over 2% of searches use one or more advanced operators	Boolean and specialized syntax are rarely used
Filter use at query level	Over 40% of queries use at least one filter	Facets/filters are more adopted than advanced syntax
Sessions beginning with filters	A little over one-third of sessions start with at least one filter	Some filters may be user-applied or library defaults
Common filters	Content type, full text, and limit to scholarly publications	Users and libraries use filters as relevance/scoping mechanisms
Suspect data in 2012 logs	Less than 0.75% of sessions and about 0.25% of queries tagged suspect	Log analysis included data-quality controls

Note. Source: Diamond, Price, and Chandrasekar (2013). Values are published aggregate indicators from anonymized Summon logs.



The log evidence confirms that discovery behaviour is often simple and front-loaded. If most clicked documents appear in the top 10 results, ranking has strong practical consequences. The very low use of fielded search and advanced operators suggests that discovery literacy cannot depend on users voluntarily applying complex search strategies; the interface and instruction must actively scaffold critical search behaviour.

6.5 Search-Engine Literacy and Algorithmic Awareness

The SEO Effect survey provides evidence on search-engine knowledge and algorithmic awareness. Although it is not a library sample, it is highly relevant because academic library users are also everyday search-engine users, and their mental models of Google-like search can transfer to discovery systems.

Table 9. Search-Engine Literacy and Algorithmic Awareness Indicators

Indicator	Reported value	Analytical meaning
Representative survey sample	N = 2,012 German internet users aged 16-69	Provides primary evidence on search-engine literacy and trust
Large-screen subsample	n = 999	Survey-device group
Small-screen subsample	n = 1,013	Survey-device group
Aware that paid advertisements exist	79%	General commercial-search awareness
Know how ads differ from organic results	29.2%	Practical result-type literacy is much lower than general awareness
Know the term search engine optimization (SEO)	8.9%	Low explicit awareness of ranking-influence terminology
Can name at least one SEO tactic correctly	14.5%	Limited procedural knowledge of ranking influence
Assume better ranking can be achieved without paying Google	43%	Awareness that ranking can be influenced outside paid ads
Assume SEO has strong or very strong impact on rankings	59.2%	Many users perceive ranking influence once SEO is considered
Perceive SEO as positive	75.2%	Users do not only interpret ranking influence negatively
Perceive SEO as negative	68.4%	Users also recognize risks or disadvantages

Note. Sources: Schultheiss and Lewandowski (2022); Lewandowski and Schultheiss (2022). Percentages are published empirical values from the representative online survey.

The key pattern is the gap between general awareness and operational literacy. While 79% knew paid advertising exists, only 29.2% knew how advertisements differ from organic results, and only 8.9% knew the term SEO. This gap supports the argument that discovery-system users may trust ranked results while lacking a precise understanding of how ranking can be influenced.



6.6 Results Mapped to the Research Questions

Table 10. Integrated Results by Research Question

Research question	Empirical result	Interpretation
RQ1: What do data show about academic-library resource use?	Students strongly used scholarly databases (84%) and OPACs (78%), but librarian consultation for assignments was only 20%.	Discovery tools and databases are primary trust gateways; human mediation is comparatively limited.
RQ2: What search behaviours are visible in discovery logs?	Summon users mostly began with simple searches; less than 6% of queries used fields, a little over 2% used advanced operators, and over 40% used filters.	Search behaviour is convenience-oriented; filters are more usable than advanced syntax.
RQ3: What do literacy data suggest about ranking awareness?	Only 29.2% knew how ads differ from organic results; only 8.9% knew SEO terminology; 59.2% assumed SEO strongly influences rankings.	Users may sense ranking influence but lack precise conceptual vocabulary.
RQ4: How do credibility and convenience shape trust?	Database credibility (78%), depth (76%), instructor expectations (74%), easy interface (65%), and one-search feature (60%) were major reasons for database use.	Trust is produced by credibility and institutional convenience together.
RQ5: What instruction is needed?	Low librarian consultation/training and simple-search behaviour point to instruction gaps.	Libraries should teach algorithmic ranking, metadata quality, database comparison, and result-list evaluation.

Note. This table synthesizes the three online empirical sources without merging their datasets.

7. Results Narrative

7.1 User Awareness of Algorithmic Bias

The available online data do not directly measure academic library users' awareness of algorithmic bias in Primo, EBSCO Discovery Service, Summon, Koha OPAC, or WorldCat Discovery. However, the search-engine literacy evidence provides a useful warning. Users show partial awareness of commercial search mediation but limited practical literacy about result types and ranking influence. The 79% awareness of paid ads contrasts sharply with the 29.2% who know how ads differ from organic results and the 8.9% who know the term SEO. This suggests that users may recognize that search is influenced, yet still lack the vocabulary and evaluative skill required to identify specific forms of influence.

7.2 Patterns of Search Behaviour

The PIL and Summon data converge on a pattern of convenience-oriented searching. Students reported heavy use of scholarly databases and OPACs, while Summon log evidence showed that most users begin with simple search rather than fielded search or advanced syntax. The use of filters in more than 40% of Summon queries indicates that facets are an important bridge between novice and advanced search. However, the limited use of fielded search and advanced operators shows that many users may not exploit the full precision of discovery tools.

7.3 Level of Digital Information Literacy

Digital information literacy appears uneven. Students value databases for credible and in-depth content, which suggests awareness of source quality. Yet the low use of librarian consultation, low participation in non-credit library



training, and low use of advanced discovery syntax point to gaps in strategic searching. The SEO Effect data reinforce this interpretation: users may use search engines confidently but show limited operational understanding of ranking influences. For academic libraries, this means that database use should not be treated as proof of discovery literacy.

7.4 Relationship Between Digital Literacy and Trust

Because the paper uses separate published aggregate sources, it cannot calculate a direct correlation or regression between digital information literacy and information trust. Nevertheless, the descriptive evidence supports an important analytical relationship. Students' database trust was linked to credibility, depth, instructor expectations, and one-search convenience. Search-engine studies show that trust can coexist with limited knowledge of ranking and commercial influence. Therefore, information literacy should aim not to reduce trust in libraries but to make trust more calibrated: users should trust scholarly systems while understanding that ranking, metadata, and coverage require critical evaluation.

7.5 Perceived Transparency of Discovery Systems

The Summon log evidence shows that a discovery system can use behavioural data to improve interface design, such as setting default results per page and identifying common query patterns. However, user-facing transparency remains a challenge. If most clicked results appear in the top 10 and users rarely use fielded or advanced search, then the default ranking and interface layout carry significant interpretive weight. Discovery systems should communicate why results are ranked, which databases are covered, when full text is missing, and how filters alter retrieval.

7.6 Differences Among User Categories

The online sources used here do not provide a comparable cross-tabulation of undergraduate students, postgraduate students, research scholars, faculty members, and librarians within one institution. Therefore, group differences are not statistically tested in this revision. A future local field study should collect role-specific survey data and run chi-square tests, ANOVA, and regression analysis to assess differences among user categories.

7.7 Qualitative Themes Derived from the Evidence

Table 11. Qualitative-Thematic Interpretation of the Online Evidence

Theme	Evidence	Interpretation
Institutional trust in scholarly databases	Students used databases because they were credible, in-depth, instructor-expected, and peer-reviewed.	Trust is attached to library-provided search spaces.
Convenience and one-search behaviour	The one-search feature was a reason for database use for 60% of respondents.	Convenience may support access but can hide coverage limits.
Simple-search dominance	Summon logs show low use of fielded search and advanced operators.	Users need interface-based and instructional support for strategic searching.
First-page/rank dependence	Most clicked Summon documents appear in the top 10 results.	Top-ranked results are likely to shape reading and citation choices.
Algorithmic literacy gap	Only 29.2% know how ads differ from organic results; 8.9% know SEO terminology.	Users may not fully understand ranking influence.
Need for librarian-led discovery literacy	Only 12% attended non-credit library training and 20% consulted librarians about assignments.	Instruction must be integrated into coursework and research workflows.

Note. Themes are derived from the descriptive evidence and literature, not from new interview transcripts.



8. Discussion

The results show that academic information trust is built through a combination of credibility cues and convenience cues. Students used scholarly databases because they believed the content was more credible than the open Internet, more detailed, and more consistent with instructor expectations. These findings align with Rieh's (2002) argument that users evaluate both information quality and cognitive authority. However, they also show that trust is not purely evaluative. Interface ease and one-search functionality were important reasons for database use, meaning that trust can be produced by convenience and institutional confidence as much as by critical evaluation.

The Summon log evidence is especially important for algorithmic bias. If most clicked documents appear in the top 10 results, ranking position has practical consequences for scholarly attention. This does not prove that top-ranked results are biased, but it shows that the consequences of ranking are substantial. When users rarely apply fielded search or advanced operators, they rely heavily on the default ranking system. This connects with Reidsma's (2019) warning that trust in library discovery may mask bias.

The search-engine literacy data deepen the interpretation. The gap between general awareness of paid ads and operational understanding of ad/organic distinction shows that users can know that search is influenced without knowing how to evaluate that influence. Transferred to academic discovery systems, this suggests that users may understand that ranking exists but not how metadata, coverage, full-text availability, local configuration, popularity signals, or recommender functions affect retrieval.

Digital information literacy should therefore be reframed as discovery literacy. It must include query formulation, controlled vocabulary, fielded search, filters, metadata interpretation, database comparison, language and coverage awareness, and reflection on rank. ACRL's view of searching as strategic exploration is useful because it treats search as an iterative scholarly process rather than a single keyword transaction (Association of College and Research Libraries, 2016).

Metadata quality and database coverage are central to trust. A record with a complete title, author, abstract, subject headings, DOI, source label, and access link is easier to evaluate than a record with weak or missing metadata. Yet metadata richness is uneven across languages, disciplines, publishers, and regions. The result is not only a cataloguing problem but also a ranking problem: ranking algorithms may interpret metadata completeness as evidence of relevance.

For librarians, the implication is practical. Instruction should include activities where users run the same query in a discovery layer, subject database, institutional repository, and Google Scholar, then compare ranking, coverage, source type, metadata completeness, peer-review indicators, and missing materials. Users should be asked to justify why a top-ranked source is trustworthy or not trustworthy. Such activities build calibrated trust rather than blind trust or generalized distrust.

For discovery-system design, the results support clearer ranking explanations, better database/source labels, visible coverage statements, transparent sort options, more useful facets, and alerts about missing metadata or limited availability. Libraries should also audit discovery performance across disciplines and languages rather than assuming that a single search box provides equal discovery for all fields.

9. Recommendations

- Integrate algorithmic-ranking awareness into library orientation, research-methods courses, and postgraduate research training.
- Teach students to compare discovery layers with subject databases, OPACs, institutional repositories, and Google Scholar rather than relying on one interface.
- Develop short discovery-literacy modules on metadata fields, peer-review filters, language coverage, citation chasing, and rank-position bias.
- Audit discovery systems using real search tasks from different disciplines and languages.
- Request vendor transparency on relevance ranking, recommender logic, metadata weighting, and personalization features.



- Improve local metadata quality in institutional repositories, especially abstracts, keywords, author identifiers, subject terms, language fields, and links.
- Embed librarians in coursework so that students encounter expert mediation before and during research assignments, not only after search failure.

10. Conclusion

This revised research paper fills the original placeholder statistics with real online primary empirical evidence while avoiding fabricated local field data. The evidence shows that academic users rely heavily on scholarly databases and OPACs, but use librarian support and formal library training far less. It also shows that discovery-system behaviour is often simple, with limited use of fielded search and advanced syntax, even though filters are more commonly used. Search-engine literacy evidence indicates that many users possess partial awareness of ranking influence but limited operational understanding of how such influence works.

The central contribution of the paper is to connect algorithmic bias, information trust, user search behaviour, and digital information literacy within one Library and Information Science framework. Trust in academic discovery is not a single attitude; it is created through the interaction of institutional authority, ranking position, metadata quality, database coverage, interface convenience, and user literacy. Therefore, academic libraries must not only provide discovery systems; they must teach users how discovery works.

The study's limitations are clear. The data sources were collected in different contexts and cannot be merged into a single inferential dataset. The Project Information Literacy data are from U.S. college students, the Summon data are published aggregate logs, and the SEO Effect survey represents German internet users rather than academic-library users. These limitations do not invalidate the evidence, but they restrict the claims to descriptive and analytical interpretation.

Future research should conduct institution-specific surveys, search-task observations, interviews, and anonymized search-log analysis in academic libraries. Such studies should test whether digital information literacy, algorithmic awareness, prior database training, discipline, language preference, and user category significantly predict information trust and critical evaluation behaviour. Multi-institutional studies in multilingual and Global South contexts are especially needed to evaluate coverage, metadata, and language bias in discovery systems.

References

- Asher, A. D., Duke, L. M., & Wilson, S. (2013). Paths of discovery: Comparing the search effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and conventional library resources. *College & Research Libraries*, 74(5), 464-488. <https://doi.org/10.5860/crl-374>
- Association of College and Research Libraries. (2016). Framework for information literacy for higher education. American Library Association. <https://www.ala.org/acrl/standards/ilframework>
- Bawden, D. (2008). Origins and concepts of digital literacy. In C. Lankshear & M. Knobel (Eds.), *Digital literacies: Concepts, policies and practices* (pp. 17-32). Peter Lang.
- Bossaller, J. S., & Sandy, H. M. (2017). Documenting the conversation: A systematic review of library discovery layers. *College & Research Libraries*, 78(5), 602-619. <https://doi.org/10.5860/crl.78.5.602>
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15, 209-227. <https://doi.org/10.1007/s10676-013-9321-6>
- Dahlen, S. P., & Hanson, K. (2017). Preference vs. authority: A comparison of student searching in a subject-specific indexing and abstracting database and a customized discovery layer. *College & Research Libraries*, 78(7), 878-897. <https://doi.org/10.5860/crl.78.7.878>
- Diamond, T., Price, S., & Chandrasekar, R. (2013). Actions speak louder than words: Analyzing large-scale query logs to improve the research experience. *The Code4Lib Journal*, 21. <https://journal.code4lib.org/articles/8693>



- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330-347. <https://doi.org/10.1145/230538.230561>
- Hanneke, R., & O'Brien, K. K. (2016). Comparison of three web-scale discovery services for health sciences research. *Journal of the Medical Library Association*, 104(2), 109-117. <https://doi.org/10.3163/1536-5050.104.2.004>
- Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring personalization of web search. *Proceedings of the 22nd International Conference on World Wide Web*, 527-538. <https://doi.org/10.1145/2488388.2488435>
- Head, A. J., & Eisenberg, M. B. (2009). Lessons learned: How college students seek information in the digital age. *Project Information Literacy*. https://projectinfoit.org/docs/information-seeking-habits/pil_info-seeking-habits_2009-12-01.pdf
- Holman, L. (2011). Millennial students' mental models of search: Implications for academic librarians and database developers. *The Journal of Academic Librarianship*, 37(1), 19-27. <https://doi.org/10.1016/j.acalib.2010.10.003>
- International Federation of Library Associations and Institutions. (2020). IFLA statement on libraries and artificial intelligence. IFLA. <https://repository.ifla.org/handle/20.500.14598/1646>
- Introna, L. D., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The Information Society*, 16(3), 169-185. <https://doi.org/10.1080/01972240050133634>
- Lewandowski, D., & Schultheiss, S. (2022). Public awareness and attitudes towards search engine optimization. *Behaviour & Information Technology*. <https://doi.org/10.1080/0144929X.2022.2056507>
- Mackey, T. P., & Jacobson, T. E. (2011). Reframing information literacy as a metaliteracy. *College & Research Libraries*, 72(1), 62-78. <https://crl.acrl.org/index.php/crl/article/view/16132/17578>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Olson, H. A. (2002). *The power to name: Locating the limits of subject representation in libraries*. Kluwer Academic Publishers.
- Pearce, A. (2019). Discovery and the disciplines: An inquiry into the role of subject databases through citation analysis. *College & Research Libraries*, 80(2), 195-214. <https://doi.org/10.5860/crl.80.2.195>
- Pulikowski, A., & Matysek, A. (2021). Searching for LIS scholarly publications: A comparison of search results from Google, Google Scholar, EDS, and LISA. *The Journal of Academic Librarianship*, 47(5), 102417. <https://doi.org/10.1016/j.acalib.2021.102417>
- Reidsma, M. (2019). *Masked by trust: Bias in library discovery*. Library Juice Press. https://scholarworks.gvsu.edu/library_books/30/
- Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, 53(2), 145-161. <https://doi.org/10.1002/asi.10017>
- Schultheiss, S., & Lewandowski, D. (2022). Data set of a representative online survey on search engines with a focus on search engine optimization (SEO): A cross-sectional study. *F1000Research*, 11, 376. <https://doi.org/10.12688/f1000research.109662.2>
- Schultheiss, S., & Lewandowski, D. (2023). Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference. *Journal of Information Science*, 49(3), 609-623. <https://doi.org/10.1177/01655515211014157>
- Schultheiss, S., Lewandowski, D., & Sünkler, S. (2022). SEO-Effekt/Online survey. OSF. <https://doi.org/10.17605/OSF.IO/PG82E>
- UNESCO. (2021). *Media and information literacy curriculum for educators and learners: Think critically, click wisely!* UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000377068>



Wang, Y., Cui, X., & Xu, S. (2018). Evaluating the impact of web-scale discovery services on scholarly content seeking. *The Journal of Academic Librarianship*, 44(5), 545-552. <https://doi.org/10.1016/j.acalib.2018.05.010>

Appendix A: Future Institution-Level Field Survey Framework

The online evidence inserted above can support a literature-based and secondary-analysis paper. For a full institution-specific empirical article, the researcher should collect local primary data from a selected university library. The following survey framework can be used without inventing values.

Table A1. Suggested Future Local Survey Items

Construct	Sample item	Scale
Digital information literacy	I compare results from more than one database or discovery tool before selecting academic sources.	5-point Likert
Algorithmic awareness	I understand that discovery systems rank search results using relevance algorithms rather than listing sources neutrally.	5-point Likert
Metadata evaluation	I examine author, journal, abstract, subject terms, date, and database source before trusting a result.	5-point Likert
Coverage awareness	I understand that a discovery system may not include all relevant publications in my discipline or language.	5-point Likert
Trust in top results	I usually trust the first page of results in the library discovery system.	5-point Likert
Training exposure	I have received formal instruction from librarians on database searching or discovery-system use.	Yes/No
Search behaviour	Which tools do you use most often for academic research?	Multiple choice

Note. These items are for future field data collection only. They are not used to generate the online statistics in this revised paper.