International Journal of Innovative Research in Management Studies (IJIRMS)
Volume 8, Issue 8, August 2026. pp.42-51.

AI-BASED WEBOMETRICS ANALYSIS: TOOLS AND TECHNIQUES

D. Daisy Kirubahari[1], Dr. Sr. M.Jaculine Mary[2], M. Prakash[3]

[1]*PG Scholar, Department of Library and Information Science, Nirmala College for Women, Coimbatore, Tamil Nadu, India – 641018. Email: mailtodaisyprakash@gmail.com*

[2]*Librarian & Head, Department of Library and Information Science, Nirmala College for Women, Coimbatore, Tamil Nadu, India – 641018. Email: joycejoseph1969@gmail.com*

[3]*Librarian, Dr. GRD Memorial Library, PSG College of Technology, Coimbatore, Tamil Nadu, India – 641004. Email: mprakashlis@gmail.com*

Abstract—*Webometrics, the quantitative study of the construction and use of Web-based information resources, has evolved considerably since its inception in the late 1990s. The exponential growth of the World Wide Web, coupled with advances in artificial intelligence (AI), has transformed how researchers and library and information science (LIS) professionals measure web presence, link structures, content quality, usage patterns and social impact of digital resources. This article examines the convergence of AI and webometrics, presenting a conceptual framework, a curated list of AI-enabled webometric tools, a step-by-step process workflow, and illustrative mock interface screens for an AI-based webometrics analysis system. Drawing on recent literature, the study discusses how machine learning, natural language processing and deep learning techniques enhance web crawling, link analysis, content and sentiment analysis, usage mining and altmetrics. The article further discusses applications for academic libraries and institutional repositories, the challenges of adopting AI in webometric research, and future directions for AI-augmented web-based research assessment.*

Keywords: *Webometrics; Artificial Intelligence; Web Impact Factor; Link Analysis; Altmetrics; Machine Learning; Library and Information Science; Web Usage Mining.*

1. Introduction

The World Wide Web has become the largest repository of scholarly, institutional and social information in human history, and measuring its structure, use and impact has become central to library and information science. The original and still dominant definition describes webometrics as "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches" (Björneborn & Ingwersen, 2004). Building on the earlier work of Almind and Ingwersen, who coined the term in 1997, Björneborn and Ingwersen (2001) further outlined webometrics as a field encompassing web-page content analysis, web-link structure analysis, web-usage analysis (including search-engine query analysis) and web-technology analysis.

For over two decades, webometric research relied largely on manual link counting, search-engine query techniques and rule-based crawlers. These approaches, while foundational, faced significant limitations: search-engine application programming interfaces changed frequently, data volumes grew far beyond what manual coding could process, and the increasingly multimedia and dynamic nature of the Web made simple hyperlink counting an incomplete indicator of web impact. The emergence of artificial intelligence encompassing machine learning (ML), natural language processing (NLP) and deep learning (DL) has opened a new phase of webometric research capable of addressing these limitations at scale. Recent systematic reviews confirm that AI techniques can enhance web crawling and data extraction, web-link analysis, web-content analysis, web-impact assessment and web-usage mining (Saeidnia et al., 2024).

For academic libraries, AI-based webometrics offers a practical pathway to evaluate institutional web visibility, benchmark digital library platforms, support National Assessment and Accreditation Council (NAAC) and National Board of Accreditation (NBA) documentation, and generate evidence for research-impact reporting. This article consolidates the conceptual foundations, tools, process workflow and interface possibilities of AI-based webometrics analysis, with the aim of providing LIS practitioners and researchers a structured reference for adopting these techniques.

The specific objectives of this article are to:

- Trace the conceptual evolution of webometrics and its convergence with artificial intelligence.
- Present a curated, categorized list of AI-enabled tools relevant to webometric analysis.
- Propose a generalized process workflow for conducting AI-based webometrics studies.
- Illustrate representative mock interface screens for an AI-based webometrics analysis system.
- Discuss applications, challenges and future directions of AI-augmented webometrics for library and information practice.

2. Literature Review

2.1 Foundations of Webometrics

The term "webometrics" was coined by Almind and Ingwersen (1997) in their study applying informetric methods to compare Denmark's web visibility with other Nordic countries. Björneborn and Ingwersen (2001) subsequently mapped the emerging field, identifying search-engine coverage, web-impact factors and content-quality assessment as key research strands, while cautioning that the analogy between citation analysis and hyperlink analysis should not be overextended. Their later work, "Toward a Basic Framework for Webometrics" (Björneborn & Ingwersen, 2004), formalized a consistent link typology and terminology, distinguishing between different Web node levels and establishing webometrics firmly within library and information science and its associated subfield, cybermetrics. Thelwall (2009) offered a complementary definition describing webometrics as the study of Web-based content using primarily quantitative methods for social-science research goals, employing techniques not specific to any one field of study.

2.2 Artificial Intelligence and Scientometric–Webometric Convergence

A growing body of literature documents how AI is reshaping scientometrics, webometrics and bibliometrics collectively. In a systematic review of 61 articles retrieved from ProQuest, IEEE Xplore, EBSCO, Web of Science and Scopus, Saeidnia et al. (2024) found that AI substantially benefits webometrics by improving web crawling and data collection, web-link analysis, web-content analysis, web-impact assessment and web-usage mining. Their review further highlighted that machine learning and data-mining techniques allow researchers to analyze user-navigation paths and interaction patterns at a scale unattainable through manual methods, while automated crawling algorithms reduce duplication and irrelevant data during collection.

Kousha and Thelwall (2024) reviewed AI applications in scholarly publishing and peer review, noting that automation is increasingly used to support journal recommendation, quality control and reviewer identification functions closely related to web-based research-impact assessment. Complementary work by Thelwall's research group on "web indicators for research evaluation" demonstrated how citations and links to academic articles from the Web, together with social-media metrics, can be systematically captured and interpreted, laying groundwork later extended by machine-learning-based citation and altmetric prediction studies. In one such study, researchers explored whether large language models such as ChatGPT could predict citation counts, Mendeley readership and social-media interaction from scientific abstracts, finding that AI-generated relevance ratings held meaningful predictive value for subsequent scholarly and social impact.

2.3 AI in Library and Information Science Practice

Within LIS practice specifically, Afjal (2023) examined the multidimensional impact of ChatGPT and generative AI, highlighting implications for reference services, information literacy and content generation in libraries. Esh and Ghosh (2024) applied contemporary AI-assisted bibliometric methods to interpret classical bibliometric laws Bradford's, Lotka's and Zipf's within the AI-in-libraries literature, subject-mapping the growing convergence of AI and library metrics. Collectively, this literature indicates that AI is not merely automating existing webometric techniques but is enabling new indicators including sentiment-aware content evaluation, predictive link-growth modelling and multi-platform altmetric aggregation that extend the scope of what webometrics can measure.

2.4 AI Techniques Mapped to Core Webometric Areas

Table 2 synthesizes the literature reviewed above by mapping the four traditional webometric research areas identified by Björneborn and Ingwersen (2001) link analysis, content analysis, usage analysis and technology analysis to the specific AI techniques now being applied within each, together with representative outputs. This mapping illustrates that AI is not replacing the conceptual categories established in classical webometrics but is instead providing more scalable and often more granular means of measuring them.

Table 2. Mapping of Classical Webometric Areas to AI Techniques

Classical Webometric Area	Representative AI Techniques	Representative AI-Derived Output
Web-link analysis	Graph neural networks; community-detection algorithms (e.g., Louvain); link-prediction models	Predicted co-link clusters; AI-adjusted Web Impact Factor
Web-content analysis	Natural language processing; transformer-based text classification; sentiment analysis	Automated topic and sentiment scores for institutional web pages
Web-usage analysis	Machine learning on clickstream and log data; anomaly detection	Predicted traffic trends; identification of unusual usage patterns
Web-technology analysis	Automated accessibility and structured-data auditing using rule-based and ML classifiers	AI-generated technical health and SEO-readiness scores

2.5 Research Gap

While reviews such as Saeidnia et al. (2024) address AI's role across scientometrics, webometrics and bibliometrics broadly, comparatively few resources synthesize AI-specific tools, a practical process workflow and interface-level illustrations focused specifically on webometrics for institutional and library contexts. This article addresses that gap by consolidating tool categories, a generalized workflow and representative interface mock-ups aimed at LIS practitioners and researchers seeking to operationalize AI-based webometric analysis.

3. Conceptual Framework of AI-Based Webometrics Analysis

Figure 1 presents a conceptual framework situating AI within the traditional webometric process. Raw web data drawn from websites, hyperlinks, search engines, social media, institutional repositories, digital libraries and altmetric platforms forms the foundation layer. This data is processed through an AI layer comprising machine learning, natural language processing and deep learning/neural-network techniques. The AI layer, in turn, feeds four AI-enhanced webometric indicator families: link analysis and web-impact factor, web-content and sentiment analysis, web-usage and traffic mining, and altmetrics/social-web impact. These indicators support downstream applications such as university and library web ranking, research-impact assessment, institutional-repository evaluation, digital-library usage insight and accreditation support, with continuous feedback refining subsequent data collection.

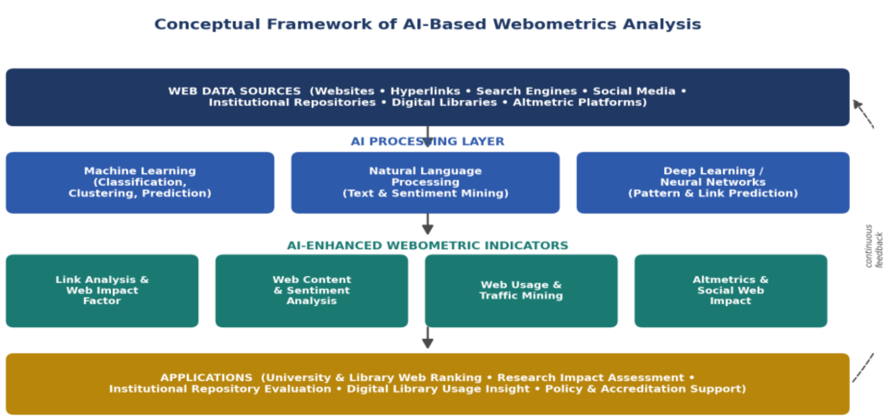

Figure 1. *Conceptual framework of AI-based webometrics analysis, showing the flow from web data sources through AI processing to webometric indicators and institutional applications.*

4. AI-Based Webometrics Tools

AI-enabled webometric tools can be broadly grouped into four categories: traditional/academic webometric tools that have incorporated AI-assisted features, AI-enhanced search-engine-optimization (SEO) and web-analytics platforms, AI–bibliometric–webometric hybrid tools that combine citation and link-based indicators, and general-purpose AI/NLP/ML development libraries used to build customized webometric pipelines. Figure 2 presents this taxonomy, and Table 1 lists representative tools with their AI capability and primary webometric use.

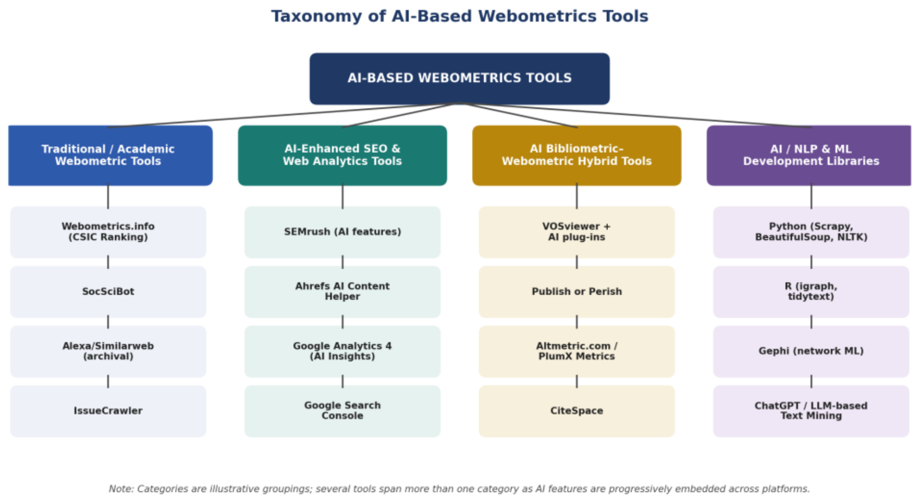

Figure 2. *Taxonomy of AI-based webometrics tools grouped into four functional categories.*

Table 1. Representative AI-Enabled Tools for Webometric Analysis

Tool / Platform	Category	AI Capability	Primary Webometric Use
Webometrics.info (CSIC Ranking Web of Universities)	Academic Webometric Ranking	Automated large-scale crawling and machine-ranking of institutional web domains	University/library web visibility and impact ranking
SocSciBot	Link Crawler & Analyzer	Rule-based crawling with statistical link-pattern detection	Academic web-link structure analysis
SEMrush (AI Overview & ContentShake AI)	AI-Enhanced SEO/Analytics	Generative-AI content briefs, AI Overview visibility tracking, intent classification	Website visibility, keyword and content-gap analysis
Ahrefs (AI Content Helper)	AI-Enhanced SEO/Analytics	NLP-based content optimization and topical-authority scoring	Backlink and content-impact analysis
Google Analytics 4 (AI Insights)	AI Web Usage Analytics	Predictive/ML-based anomaly detection and audience insights	Web usage and visitor-behaviour mining

Tool / Platform	Category	AI Capability	Primary Webometric Use
VOSviewer (with AI-assisted term extraction)	Bibliometric-Webometric Hybrid	NLP term/keyword clustering for co-occurrence mapping	Visualizing web/citation network clusters
Altmetric.com / PlumX Metrics	Altmetrics Platform	ML-based aggregation and classification of social-web mentions	Tracking social and policy impact of web-hosted research
Gephi (with ML/Graph plug-ins)	Network Visualization	Graph-neural and community-detection algorithms (e.g., Louvain)	Visualizing hyperlink and co-citation networks
Python (Scrapy, BeautifulSoup, NLTK, spaCy)	AI/NLP Development Toolkit	Custom machine learning, NLP pipelines and automated crawlers	Bespoke webometric data extraction and text mining
ChatGPT / other LLM-based assistants	Generative AI / LLM Tool	Large-language-model summarization, classification and query answering	Rapid content analysis, abstracting and reporting support

The list in Table 1 is illustrative rather than exhaustive. Industry data indicate that AI capabilities are being embedded rapidly across mainstream SEO and analytics platforms for example, Ahrefs introduced its AI Content Helper in September 2024 with subsequent enhancements, while Semrush consolidated its AI features into a unified suite and, together with Ahrefs, launched Model Context Protocol (MCP) servers in 2025 enabling AI agents to query SEO data through natural-language interfaces. For academic and institutional purposes, however, purpose-built webometric ranking systems and bibliometric-visualization tools such as VOSviewer and Gephi remain central, particularly when combined with AI-assisted clustering and graph-neural link-prediction algorithms.

5. Process Workflow for AI-Based Webometrics Analysis

Figure 3 illustrates a generalized ten-step process workflow suitable for conducting an AI-based webometrics study, whether for an individual institution, a group of universities, or a digital-library platform. The workflow begins with defining the research scope and target web domains, followed by AI-driven crawling and automated data extraction. A data-sufficiency decision point ensures that inadequate or irrelevant data triggers a refinement loop back to the crawling stage before proceeding to data cleaning, NLP-based content and sentiment analysis, machine-learning-based link and network analysis, AI-enhanced usage and altmetric mining, AI-assisted visualization, and finally interpretation and decision-support reporting.

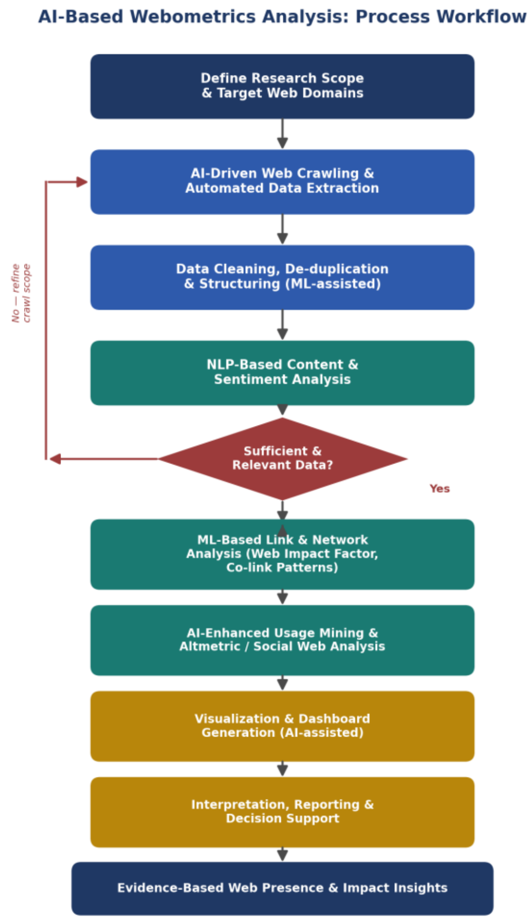

Figure 3. *Process workflow for conducting an AI-based webometrics analysis, from scope definition to evidence-based reporting.*

Each stage of this workflow can draw on the tool categories identified in Table 1. For instance, AI-driven crawling may use Python-based frameworks such as Scrapy in combination with machine-learning classifiers to prioritize relevant pages; NLP-based content analysis may apply libraries such as spaCy or NLTK, or large-language-model assistants, to classify and summarize page content; and network analysis may apply community-detection algorithms available in Gephi or VOSviewer to reveal clusters of related institutional web resources.

6. Mock Interface Screens for an AI-Based Webometrics Analysis System

To illustrate how the conceptual framework and workflow translate into a usable analytical environment, this section presents two representative mock interface screens for a hypothetical "AI-Webometrics Analyzer" system designed for institutional and library use.

6.1 AI Webometrics Dashboard

Figure 4 shows a mock dashboard summarizing key AI-derived webometric indicators an AI-adjusted web-impact factor, total inbound links detected, an average NLP-derived sentiment score, and an AI content-quality index alongside trend and distribution charts and a table of AI-flagged high-impact web pages with recommended actions. Such a dashboard enables librarians and administrators to monitor institutional web visibility at a glance.

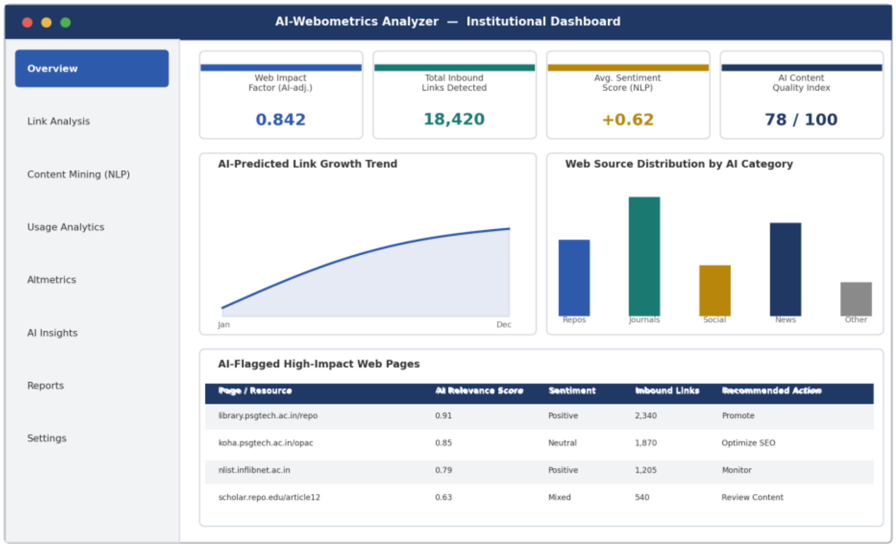

Figure 4. *Mock screen of an AI-based webometrics analysis dashboard showing key performance indicators, trend charts and AI-flagged high-impact pages.*

6.2 AI Link and Co-citation Network Explorer

Figure 5 presents a mock network-exploration screen in which AI clustering algorithms (for example, Louvain community detection combined with graph-neural-network link prediction) group institutional web resources into clusters such as repository pages, journal and database platforms, and social or repository mentions while dashed bridge links indicate AI-identified cross-cluster connections warranting further investigation.

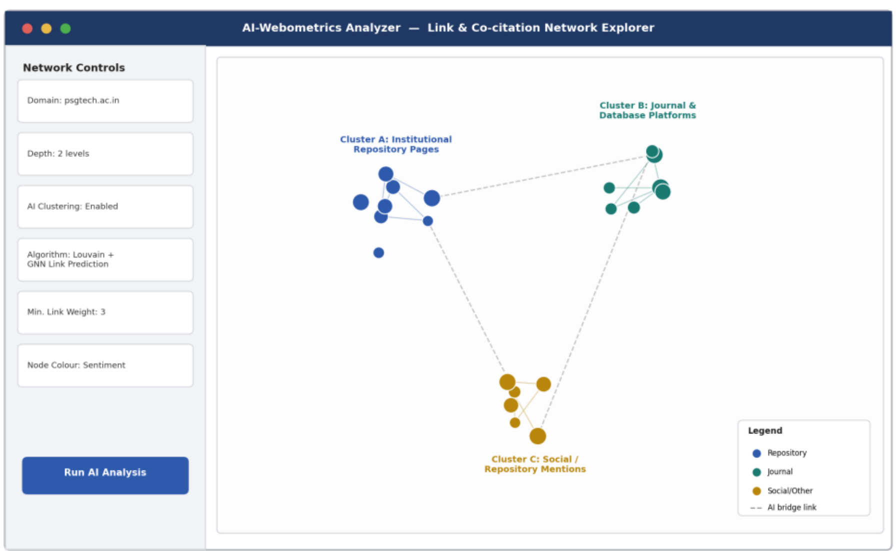

Figure 5. *Mock screen of an AI-based link and co-citation network explorer showing clustered web resources and AI-predicted bridge connections.*

7. Discussion: Traditional versus AI-Based Webometrics

It is useful to situate AI-based webometrics against the traditional, manually intensive practices that dominated the field for over two decades. Table 3 contrasts the two approaches across several dimensions relevant to library and information science practice, including data-collection scale, indicator granularity, turnaround time and the level of technical skill required. The comparison suggests that AI does not so much replace the conceptual apparatus developed by Björneborn and Ingwersen (2001, 2004) and Thelwall (2009) as it re-instruments it automating labour-intensive stages such as link counting and content coding while introducing genuinely new capabilities such as sentiment-aware content evaluation and predictive link-growth modelling that were impractical under manual methods.

Table 3. Traditional Webometrics versus AI-Based Webometrics

Dimension	Traditional Webometrics	AI-Based Webometrics
Data collection	Manual link counting; search-engine query techniques; small-scale rule-based crawlers	Automated, large-scale AI-driven crawling with intelligent page-relevance filtering
Content evaluation	Manual coding of page content and topic categories	NLP-based automatic classification, topic modelling and sentiment analysis
Link/network analysis	Manual or semi-automated link counting; basic network diagrams	Graph-neural-network link prediction; automated community detection
Usage analysis	Server-log review with limited statistical summary	Machine-learning-based pattern detection and anomaly/trend prediction
Turnaround time	Weeks to months for a moderate-sized study	Hours to days for comparable or larger-scale studies
Technical skill required	Bibliometric/webometric domain knowledge; basic spreadsheet or scripting skills	Domain knowledge plus familiarity with AI/ML tools, APIs and data pipelines
Key limitation	Difficult to scale; sensitive to manual coding bias and fatigue	Sensitive to algorithmic bias, data-access restrictions and model interpretability

This comparison also clarifies why AI-based webometrics should be understood as complementary rather than a wholesale substitute for established methods. Well-validated indicators such as the Web Impact Factor and established link-typology terminology remain conceptually important; AI primarily changes how efficiently and at what scale these indicators, and newer AI-derived indicators, can be computed and interpreted (Saeidnia et al., 2024). For library and information science professionals, this implies that domain expertise in webometric theory remains essential even as the underlying computational methods become increasingly automated.

8. Applications and Benefits

AI-based webometrics offers several practical benefits for libraries, universities and research institutions, including:

- Web visibility ranking and benchmarking of institutional repositories, library catalogues (e.g., Koha-based OPACs) and digital-library platforms against peer institutions.
- Evidence generation for accreditation processes such as NAAC and NBA, where digital infrastructure visibility and usage indicators are increasingly relevant.
- Automated, large-scale content and sentiment analysis of institutional web presence to identify pages requiring improvement.
- Predictive analytics for link growth and web-traffic trends, supporting proactive digital-library planning.
- Integrated altmetric and social-web impact tracking for research outputs hosted in institutional repositories.
- Time and labour savings compared with manual link-counting and content-coding methods used in earlier webometric studies.

As an illustration, an academic library operating a Koha-based integrated library management system alongside an institutional repository could apply the workflow in Figure 3 to crawl its own web domain and those of peer institutions, apply NLP-based content scoring to repository landing pages, and use graph-based clustering to identify which external journal platforms, aggregators (e.g., N-LIST, Shodhganga-type repositories) and social platforms most frequently link to its collections. The resulting dashboard-style output, similar to Figure 4, could then feed directly into NAAC/NBA self-study documentation as quantitative evidence of digital-library visibility and reach.

9. Challenges and Limitations

Despite its promise, the adoption of AI in webometric research faces several challenges:

- Data access constraints: frequent changes to search-engine and social-platform application programming interfaces can disrupt AI-driven data-collection pipelines.
- Algorithmic bias: machine-learning models trained on skewed or non-representative web data may produce biased impact or ranking estimates.
- Interpretability: complex deep-learning and graph-neural models can behave as “black boxes,” complicating the interpretation of AI-derived webometric indicators for policy or accreditation purposes.
- Zero-click and AI-summarized search: the rise of AI-generated search overviews is altering traditional click-through and traffic patterns, requiring webometric indicators to be recalibrated for an AI-mediated search environment.
- Resource and skill requirements: implementing custom AI-based crawling and NLP pipelines requires technical expertise not always available within library and information science units.
- Ethical and privacy considerations: AI-based usage-mining techniques must be applied in a manner consistent with user-privacy expectations and institutional data-governance policies.

10. Future Directions

Future research and practice in AI-based webometrics is likely to move in several directions. First, large-language-model-based assistants are expected to play a growing role in automating content summarization, sentiment classification and even natural-language querying of webometric datasets, reducing the technical barrier for LIS practitioners. Second, as AI-generated search overviews and “zero-click” results become more prevalent, webometric indicators will need to expand beyond traditional click and link counts to capture visibility within AI-mediated answer engines. Third, graph-neural-network-based link-prediction methods offer potential for forecasting emerging institutional web relationships before they are fully established, supporting proactive digital-library strategy. Finally, greater standardization of AI-based webometric **indicatorsakin** to the standardization achieved for traditional bibliometric indices would strengthen their credibility for institutional ranking, accreditation and research-assessment purposes.

11. Conclusion

Webometrics has matured from a niche extension of bibliometrics into an essential tool for understanding institutional and scholarly web presence. The integration of artificial intelligence through machine learning, natural language processing and deep learning is extending the reach, scale and sophistication of webometric analysis, enabling more automated crawling, richer content and sentiment analysis, and predictive link and usage modelling than traditional manual techniques allowed. This article has presented a conceptual framework, a categorized list of AI-enabled tools, a generalized process workflow and illustrative mock interface screens to guide library and information science practitioners and researchers in adopting AI-based webometrics analysis. As AI capabilities continue to be embedded across both academic and commercial web-analytics platforms, libraries that build the requisite skills and infrastructure will be well positioned to generate timely, evidence-based insight into their institutional web presence and impact.

References

[1] Afjal, M. (2023). ChatGPT and the AI revolution: A comprehensive investigation of its multidimensional impact and potential. Library Hi Tech, ahead-of-print. https://doi.org/10.1108/LHT-07-2023-0322

[2] Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to “webometrics.” Journal of Documentation, 53(4), 404–426.

[3] Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. Scientometrics, 50(1), 65–82. https://doi.org/10.1023/A:1005642218907

[4] Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. Journal of the American Society for Information Science and Technology, 55(14), 1216–1227. https://doi.org/10.1002/asi.20077

[5] de Winter, J. (2024). Can ChatGPT be used to predict citation counts, readership, and social media interaction? An exploration among 2222 scientific abstracts. Scientometrics, 129, 2469–2487. https://doi.org/10.1007/s11192-024-04939-y

[6] Esh, M., & Ghosh, S. (2024). AI in libraries: Subject mapping and interpreting bibliometric laws with contemporary metrics. The Serials Librarian, 85(5–6), 99–116. https://doi.org/10.1080/0361526X.2024.2444288

[7] Kousha, K., & Thelwall, M. (2024). Artificial intelligence to support publishing and peer review: A summary and review. Learned Publishing, 37(1), 4–12. https://doi.org/10.1002/leap.1570

[8] Saeidnia, H. R., Hosseini, E., Abdoli, S., & Ausloos, M. (2024). Unleashing the power of AI: A systematic review of cutting-edge techniques in AI-enhanced scientometrics, webometrics and bibliometrics. Library Hi Tech, 43(4–5), 1292–1314. https://doi.org/10.1108/LHT-10-2023-0514

[9] Thelwall, M. (2009). Introduction to webometrics: Quantitative Web research for the social sciences. Morgan & Claypool Publishers.

*****