Chapter 12 / Capítulo 12

Applied bibliometrics. From data to publication (English Edition)
ISBN: 978-9915-9680-7-0
DOI: 10.62486/978-9915-9680-7-0.ch12

Emerging Trends / Tendencias Emergentes

The contemporary bibliometric landscape is undergoing a radical transformation, driven by the convergence of disruptive technologies that are substantially expanding the frontiers of what can be analyzed and interpreted. These technological innovations not only optimize established methodologies but also fundamentally redefine the research questions that can be asked and answered through quantitative analysis of the scientific literature. From generative artificial intelligence to distributed ledger technologies, emerging trends promise to overcome historical limitations in the field while introducing new ethical and methodological challenges that the bibliometric community must address with rigor and foresight. This chapter critically examines these transformative trends, assessing both their analytical potential and their implications for the future of scientific evaluation.

12.1. Generative AI: GPT in scientific text mining

Large-scale language models such as GPT represent a disruptive innovation in bibliometric analysis by enabling the contextual understanding of scientific texts at scale. Unlike traditional text mining methods based on term frequency or co-occurrence analysis, these models capture complex semantic relationships and interpretive nuances that previously required specialized human intervention.

This capability substantially transforms bibliometric content analysis, enabling everything from the automatic classification of articles according to complex epistemological dimensions to the identification of argumentative and rhetorical patterns that transcend the mere presence or absence of specific terms. The application of these models to entire bibliographic corpora opens up unprecedented analytical possibilities for understanding the discursive evolution of scientific fields.

The implementation of generative AI in bibliometrics enables the automated extraction of analytical dimensions that are traditionally inaccessible through conventional quantitative methods. Models can automatically identify key theoretical contributions, methodological innovations, and empirical findings in scientific publications, classifying articles according to complex taxonomies without requiring predefined terminological dictionaries. This analytical flexibility is particularly valuable for studying interdisciplinary fields where specialized vocabularies evolve rapidly and where intellectual contributions transcend established disciplinary categorizations. The ability of these systems to generate analytical summaries synthesizing findings from multiple publications further amplifies their usefulness for bibliometric reviews of large volumes of literature.

The analysis of thematic trends using generative AI significantly overcomes the limitations of keyword-based methods by capturing conceptual and semantic evolutions beyond mere terminological frequency. Models can track how certain concepts acquire different meanings in different temporal or disciplinary contexts, identifying processes of theoretical recontextualization or conceptual appropriation between fields. This ability to map the semantic drift of scientific ideas provides unique insights into the intellectual dynamics of research fields, revealing how conceptual frameworks are transformed through their circulation alongside different epistemic communities.

This diachronic analysis of semantic evolution constitutes a fundamental methodological contribution of generative AI to contemporary bibliometrics.

The automatic generation of research hypotheses is another transformative application in which AI systems analyze patterns in scientific literature to identify knowledge gaps and promising research opportunities. By processing entire document corpora, these systems can detect non-obvious connections between seemingly unrelated findings, suggest innovative combinations of methodological approaches, or identify under-explored research problems with high impact potential. This capacity for creative synthesis across traditionally separate domains of knowledge and generative AI positions it as a valuable tool for strategic research planning and the identification of emerging scientific frontiers.

The evaluation of argumentative quality and methodological soundness using generative AI introduces qualitative dimensions previously inaccessible at the bibliometric scale. Models can systematically analyze the argumentative structure of publications, evaluate the adequacy of research questions and the methods used, and identify recurring methodological limitations within specific literatures.

This ability to assess quality dimensions beyond citations offers promising alternatives to scientific evaluation systems that transcend traditional impact metrics. However, it introduces significant challenges of validation and algorithmic transparency.

The implementation of generative AI in bibliometrics faces significant challenges, including methodological transparency, biases in training data, and the reproducibility of analyses. Language models exhibit a tendency to generate factual hallucinations, reproduce biases present in their training data, and exhibit inconsistent responses across different requests. These challenges require the development of rigorous validation protocols, comprehensive documentation of parameters and prompts used, and the implementation of explainability mechanisms that allow for understanding the rationale behind automatically generated classifications and analyses.

The integration of generative AI with traditional bibliometric methods establishes a hybrid analytical paradigm that combines the strengths of both approaches. Established citation and network analysis techniques provide quantitative validation for AI-generated qualitative insights, while advanced semantic analysis contextualizes and enriches the interpretation of structural patterns identified through traditional methods.

This methodological integration produces more robust and nuanced bibliometric analyses that leverage the best of both analytical paradigms, setting new standards for a comprehensive understanding of scientific dynamics.

The future development of generative AI applied to bibliometrics will likely be oriented toward models specialized in specific scientific domains, trained on comprehensive disciplinary corpora, and capable of understanding the rhetorical and epistemological conventions particular to each field. The evolution toward multimodal systems that integrate the analysis of text, images, data, and code in scientific publications will further expand the frontiers of bibliometric analysis, enabling a more comprehensive understanding of contemporary scientific communication. This progressive specialization conditions current limitations of general models while maximizing their usefulness for specific bibliometric applications.

12.2. Blockchain for decentralized metrics

Blockchain technology is emerging as a transformative solution to address fundamental challenges of transparency, trust, and decentralization in academic metrics systems. Unlike traditional centralized databases, distributed ledgers enable the creation of incorruptible metric systems in which every academic transaction—citations, reviews, contributions—is recorded in an immutable, verifiable manner by any participant in the network.

This decentralized architecture eliminates single points of failure and reduces dependence on institutional intermediaries, laying the foundation for a more resilient and reliable metrics ecosystem. The application of blockchain to bibliometrics represents a paradigm shift toward more transparent, tamper-resistant scientific evaluation systems.

The implementation of blockchain allows for the creation of decentralized records of academic contributions that capture the entire research process, not just its final products. Each contribution, from initial conceptualization and methodological design to data analysis and writing, can be recorded with an unalterable timestamp on the blockchain, creating auditable traces of authorship and intellectual contribution. This approach solves chronic problems of authorship attribution and recognition of non-traditional contributions, particularly valuable in collaborative research where multiple researchers contribute in different capacities throughout the research cycle. The granularity of contribution recording enables fairer, more accurate metrics that reflect the complex, collaborative nature of contemporary science.

Blockchain-based metrics systems introduce innovative mechanisms for verifying citations and preventing metric manipulation.

Each citation can be recorded as a transaction on the chain, creating a transparent and immutable history of intellectual influence relationships.

This distributed record enables detection of circular citation patterns, coordinated self-citation networks, and other manipulative practices that distort traditional metrics. Decentralized verification through network consensus eliminates reliance on centralized databases whose processing and cleaning algorithms often operate as black boxes inaccessible to end users.

Academic tokens and decentralized reputation systems represent another disruptive application of blockchain in bibliometrics. These systems allow researchers to accumulate reputation based on verified contributions recorded on the chain, creating portable academic capital that is independent of specific institutions. Tokenization mechanisms can align individual incentives with collective goals of the scientific system, rewarding valuable academic behaviors such as rigorous peer review, sharing research data, or replicating previous studies. This approach fundamentally transforms academic reward systems by making values traditionally implicit in scientific culture explicit and quantifiable.

Decentralized management of academic identities through blockchain solves persistent problems of disambiguation and portability of researcher profiles. Each researcher can maintain a self-sovereign identity registered on the chain, consolidating their contributions across different institutions, platforms, and time periods. This persistent, vendor-independent identity facilitates academic mobility and eliminates the fragmentation of metrics that occurs when researchers change institutional affiliations. Integration with existing ORCID systems and other academic identifiers creates a more robust, user-centric identity ecosystem.

Smart contracts enable the automation of scientific evaluation processes based on metrics verified on the chain. These self-executing protocols can implement transparent evaluation criteria and automate processes such as peer review, funding allocation, or academic promotions based on objective metrics recorded on the blockchain.

Automation via smart contracts reduces the administrative burden on scientific evaluation while increasing the consistency and transparency of decision-making. However, this automation requires careful design to avoid encoding existing biases or creating overly rigid systems that fail to capture critical qualitative dimensions of scientific excellence.

The implementation of decentralized metrics faces significant challenges of scalability, adoption, and governance. Recording every academic transaction on the chain generates massive volumes of data that must be processed efficiently. In contrast, adoption by diverse academic communities requires overcoming institutional inertia and standardizing protocols across different disciplines and regions. Governance models for these decentralized systems must balance community participation with decision-making efficiency, avoiding both capture by particular interests and paralysis by excessive deliberation.

These practical challenges require careful attention during the design and implementation of blockchain-based bibliometric systems.

The integration of blockchain with other emerging technologies, such as artificial intelligence and the Internet of Things, creates even more sophisticated and comprehensive metric ecosystems. AI systems can analyze immutable data recorded on the chain to generate more reliable insights, while IoT devices can automate the recording of research activities in laboratories and field environments.

This technological convergence enhances the capabilities of each technology while mitigating its specific limitations, creating more robust, transparent metric infrastructures aligned with actual scientific research practices across different domains and methodologies.

12.3. Image and multimodal data analysis

Multimodal image and data analysis represents an emerging frontier in bibliometrics that significantly expands the scope of what can be quantified in scientific communication. Traditionally, bibliometric studies have focused predominantly on text analysis, neglecting the rich visual content that accompanies contemporary academic publications. Images, graphs, diagrams, and visualizations are essential components of modern scientific discourse, conveying complex information that often cannot be adequately expressed through text alone. The ability to systematically analyze these visual elements opens up new dimensions for understanding how knowledge is constructed and communicated across different scientific disciplines.

Computer vision and deep learning techniques enable the extraction of meaningful information from visual elements in scientific publications at scale. Algorithms can automatically identify and classify images across different types, from statistical graphs and flowcharts to microphotographs and three-dimensional visualizations, establishing usage patterns by specific disciplines, methodologies, or time periods.

This analytical capability transforms images from mere illustrations into quantifiable data that reveal representational preferences and standards, visualization standards, and evolutions in visual communication practices within different scientific communities. The automated analysis of these visual elements complements traditional textual approaches, offering a more comprehensive perspective on visual rhetoric in science.

Multimodal analysis integrates information from different formats—text, images, tables, equations—to generate richer, more contextualized understandings of scientific content. Rather than analyzing each modality separately, multimodal approaches capture the interrelationships between different forms of knowledge representation, revealing how researchers combine diverse semiotic resources to construct persuasive scientific arguments.

This integration allows, for example, analysis of how textual descriptions relate to corresponding visualizations, or how mathematical equations are complemented by narrative explanations and graphic representations. Multimodal bibliometrics thus captures the essentially hybrid nature of contemporary scientific communication.

The detection of methodological trends through image analysis is a particularly valuable application for scientific mapping. Certain types of images serve as reliable indicators of specific methodological approaches: electron micrographs suggest certain laboratory techniques, while circuit diagrams point to particular experimental approaches in engineering. Diachronic analysis of the prevalence of different image types can reveal the adoption, consolidation, or decline of research methodologies, providing insights into the evolution of scientific practices that complement analyses based on textual terminology. This ability to track methodologies through their visual traces represents a unique contribution of multimodal analysis to bibliometrics.

The evaluation of visual quality and clarity in scientific publications emerges as another promising application of multimodal analysis. Algorithms can evaluate aspects such as image resolution, appropriateness of scale, label clarity, and the communicative effectiveness of different types of visualizations. This automated assessment of visual quality enables large-scale studies of the evolution of scientific visualization standards, the identification of best practices in visual communication, and the analysis of how visual quality correlates with scientific impact as measured by citations and other indicators.

This traditionally subjective qualitative dimension is thus transformed into an object of systematic quantitative analysis.

Multimodal data analysis faces significant methodological challenges in standardization, interpretation, and validation. The enormous diversity of visual formats in scientific disciplines complicates the development of universally applicable taxonomies and algorithms. Furthermore, interpreting visual elements often requires specialized domain knowledge, making it difficult to fully automate the analysis without losing crucial disciplinary nuances. Validating multimodal analysis results requires innovative strategies that combine automatic evaluation with human expert verification, establishing reliability standards for this new bibliometric frontier.

The integration of multimodal analysis with other emerging trends, such as generative AI and blockchain, creates compelling synergies. Multimodal language models can generate textual descriptions of visual content or create visualizations from textual descriptions, facilitating the accessibility and searchability of multimodal scientific content. Simultaneously, blockchain can provide the infrastructure to record and verify the provenance of multimodal images and data, addressing growing concerns about the manipulation of scientific images.

These technological integrations amplify the transformative potential of multimodal analysis in bibliometrics.

The future development of multimodal bibliometric analysis will evolve toward systems capable of capturing increasingly sophisticated dimensions of integrated scientific communication. The ability to analyze not only static images but also interactive visualizations, scientific videos, and immersive virtual reality representations will further expand the frontiers of what can be analyzed.

12.4. Ethics and transparency in the age of AI

The implementation of artificial intelligence systems in bibliometrics raises fundamental ethical challenges that require robust governance frameworks. The algorithmic opacity characteristic of many contemporary models creates significant risks of reproducing and amplifying historical biases in scientific evaluation, perpetuating structural inequalities. This lack of transparency undermines the legitimacy of AI-based bibliometric evaluation systems, requiring methodological approaches that prioritize the explainability and auditability of decision-making processes to maintain the academic community's trust in these emerging tools.

Ensuring methodological transparency is an essential pillar, supported by comprehensive documentation of training data, model architectures, and validation procedures. Metadata should include detailed information on the demographic and geographic composition of the datasets used, enabling identification of representational gaps that may distort results. Researchers have an ethical responsibility to disclose both the known capabilities and limitations of their systems, avoiding undue extrapolations beyond the original development and validation conditions, which facilitates critical peer review.

The protection of personal data emerges as a critical consideration when systems process sensitive information from researchers, including career trajectories and collaboration networks. Ethical frameworks must ensure regulatory compliance, implement minimization principles in data collection, and establish clear informed consent protocols.

These safeguards are critical in evaluative contexts where results can affect career opportunities and resource allocation, requiring careful balances between analytical utility and the protection of individual privacy in the research ecosystem.

Accountability for algorithmic decisions represents another central challenge, requiring transparent chains of responsibility when automated systems generate rankings or evaluate scientific merit. Effective appeal mechanisms should be established for cases where researchers or institutions consider evaluations to be unfair, complemented by ethical oversight committees that incorporate disciplinary and geographic diversity.

This distributed governance structure allows for the identification of operational risks and the establishment of protocols for responsible use that maintain institutional trust in these emerging bibliometric evaluation systems.

Proactively mitigating algorithmic biases requires systematic strategies that include developing specialized techniques, conducting regular equity audits, and diversifying development teams. Continuous assessment of the differential impact across demographic groups, institutions, and geographic regions allows for the identification and correction of emerging disparities before they become entrenched as structural injustices.

This preventive approach is essential for building bibliometric systems that reflect the values of equity and inclusion inherent in contemporary scientific enterprise in its epistemological and methodological diversity.

Ecological sustainability completes the ethical picture by accounting for the massive computational requirements that generate significant carbon footprints during model training and deployment. Developers must prioritize energy efficiency, select balanced architectures, and make environmental costs transparent, aligning technological innovation with the imperatives of environmental responsibility that characterize contemporary science. This ecological awareness represents an essential dimension of ethics applied to the development of bibliometric tools powered by artificial intelligence in the current context of the global climate crisis.

Participatory governance involves multiple actors—researchers, managers, institutional representatives—in the design and oversight of systems, ensuring that diverse perspectives inform their development. Establishing open deliberative processes on embedded values, trade-offs between objectives, and accountability standards builds social consensus around the appropriate use of these technologies. This inclusiveness mitigates the risks of technocratic imposition and ensures that systems reflect the plural values of the global scientific community in its entirety.

Training in digital ethics develops critical users capable of interpreting algorithmic results through programs that emphasize human judgment as an essential complement to automated evaluations. Cultivating informed skepticism toward counterintuitive or potentially biased results prevents uncritical applications, while strengthening community capacities to reap benefits and mitigate risks. This literacy is a necessary investment in the human capital that will use these tools, ensuring their responsible implementation in the diverse institutional and disciplinary contexts where they will be applied.

Recap

- Open science promotes transparency and reproducibility.
- The OpenAlex and Crossref platforms democratize access to metadata.
- Artificial intelligence facilitates trend analysis and prediction.
- Altmetrics extend impact measurement beyond citations.
- The use of interactive dashboards improves dynamic visualization.
- Algorithmic transparency is fundamental in AI-based bibliometrics.
- Big data enables real-time analysis of scientific output.
- Equity indicators assess diversity across gender, language, and geography.
- Integrating patents, publications, and funding improves innovation assessment.
- Reproducible tools such as R and Python facilitate replication.
- Risks: algorithmic manipulation and metric gaming.
- Manifests such as DORA and Leiden promote responsible evaluation.
- Open repositories strengthen collaborative science.
- Data interoperability (DOI, ORCID) improves traceability.
- Alternative metrics reflect social and media impact.
- Web dashboards enable continuous monitoring of indicators.
- Metric reports are evolving toward narrative visualizations.
- AI ethics becomes a priority in research.
- Predictive bibliometrics combines algorithms and trend monitoring.
- Training in analysis tools will become essential in the future.

Self-assessment questions

1. What changes does AI bring to bibliometric analysis?
2. What is the difference between altmetrics and traditional citations?
3. What open sources are relevant for new studies?
4. What are the main ethical risks of AI in scientific evaluation?

5. What does the DORA Declaration promote?
6. Why is data interoperability important?
7. What advantages do interactive panels offer?
8. What does the concept of responsible evaluation entail?
9. What tools are key to reproducible analysis?
10. How can bibliometrics predict emerging trends?

BIBLIOGRAPHY

1. Sugimoto CR, Larivière V. Measuring Research: What Everyone Needs to Know. Oxford: Oxford University Press; 2018. ISBN: 9780190640125.

2. Thelwall M. Web Indicators for Research Evaluation: A Practical Guide. San Rafael (CA): Morgan & Claypool; 2016. DOI: 10.2200/S00733ED1V01Y201602ICR048.

3. Gingras Y. Bibliometrics and Research Evaluation: Uses and Abuses. Cambridge (MA): MIT Press; 2016. ISBN: 9780262337663.

4. Wilsdon J, et al. The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management. London: HEFCE; 2015. ISBN: 9780992492124.