

Del SEO académico a la optimización académica para modelos de lenguaje (ALO): un marco conceptual para la visibilidad académica en la era de la inteligencia artificial generativa

Dr. **Carlos Lopezosa**

PrimeTime

Jornada de transferencia de los resultados del proyecto

Valencia, 26 de junio 2026



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Introducción

- 10 prompts sobre evaluación de la ciencia
- 10 herramientas de IA
- Toma de datos (octubre y noviembre-2025)

Unraveling the Ai2 Asta Scholarly Research Assistant Citation System

Descifrando el sistema de citas del asistente de investigación académica Ai2 Asta

Enrique Orduña-Malea; Carlos Lopezosa



Enrique Orduña-Malea
Universitat Politècnica de València, Spain
The iMetrics Lab
<https://orcid.org/0000-0002-1989-8477>
enorma@upv.es



Carlos Lopezosa
Universitat de Barcelona, Spain
Dep. de Biblioteconomia, Doc. i
Comunicació Audiovisual
<https://orcid.org/0000-0001-8619-2194>
lopezosa@ub.edu

Cómo citar este artículo

Orduña-Malea, E., & Lopezosa, C. (2025). Unraveling the Ai2 Asta Scholarly Research Assistant Citation System. *Revista Panamericana de Comunicación*, 7(2), 3675. <https://doi.org/10.21555/rpc.v7i2.3675>

Recibido: 25 - 11 - 2025

Aceptado: 04 - 12 - 2025

Publicado en línea: 06-12-2025

Abstract

Despite the growing integration of Deep Research tools into academic workflows, empirical evidence on the operation, stability, and potential biases of their citation systems remains scarce. This study addresses this gap by evaluating the intensity, consistency, and bibliographic characteristics of references cited in the literature reports generated by Ai2 Asta, with the aim of understanding how its citation system operates and assessing its implications for scholarly communication. To this end, ten domain-specific queries were submitted to Asta's Summarise Literature feature, and two independent rounds of data collection were conducted. From each report, in-text citations, cited references, as well as other metrics related to the response process were extracted and examined. The results reveal high citation intensity, with reports integrating numerous in-text citations grounded in retrieved evidence and a diverse yet concentrated set of venues. However, notable instability is observed in the composition of cited references across identical queries, alongside a lack of concordance between retrieved documents and those ultimately cited, suggesting additional opaque selection mechanisms during report generation. These findings indicate that, while Ai2 Asta produces well-structured and quality reports, its instability and opacity in the citation process pose challenges in quantitative science studies due to their lack of reproducibility and transparency. Despite the restricted number of queries and disciplinary scope, the

ALO

- **ALO (Academic Large Language Model Optimization)** puede definirse como el proceso de diseñar y publicar trabajos científicos de forma que maximicen su probabilidad de ser **recuperados, comprendidos, citados** y utilizados como evidencia por sistemas de inteligencia artificial generativa y buscadores académicos basados en grandes modelos de lenguaje.

Herramientas de IA utilizadas

- ChatGPT
- Gemini
- Perplexity
- Elicit
- Asta
- Scopus AI
- WoS AI
- Scispace
- Epsilon

What is the current state of the Webometrics (also known as Cybermetrics) discipline in 2025, considering its advantages, limitations, current applications, and main lines of research?

S'està cercant Webometrics discipline webometrics cybermetrics research trends 2025 bibliometrics altmetr

Instrumento de evaluación

- Referencia
- Título del artículo
- Incluye alguna palabra del prompt
- Abstract del artículo incluye alguna palabra del prompt
- las kw del artículo incluye alguna palabra del prompt
- artículo indexado en scopus
- artículo indexado en wos
- número de citas recibidas en scholar
- número de citas recibidas en scopus
- número de citas recibidas en wos
- fuente/editorial de la revista
- año de publicación
- cuartil de la revista en el año de publicación
- número de autores
- índice h de autor 1 (en scholar)
- índice h de autor 2 (en scholar)
- índice h de autor 3 (en scholar)
- índice h de autor 4 (en scholar)

Resultados preliminares: ChatGPT

- 13 resultados recuperados entre el año 2008 y 2025 (8 documentos académicos y 5 de Wikipedia, páginas institucionales, rankings y sitios web).
- 100% de los documentos académicos contienen términos del prompt en el abstract.
- 88% de los documentos académicos contienen términos del prompt en el título (7/8).
- 75% de los documentos académicos contienen términos del prompt en las palabras clave (6/8).
- 25% de los documentos académicos están indexados en Scopus (2/8), ninguno de ellos en WoS
- Promedio de citas en Google Scholar (documentos con datos disponibles): ≈ 59 citas, aunque muy influido por un artículo clásico con 389 citas.
- Predominio de fuentes researchgate, arxiv y preprints.org

What is Webometrics / Cybermetrics

- The term *webometrics* was introduced by Tomas Almind & Peter Ingwersen in 1997 to denote "the study of quantitative aspects of the construction and use of information resources, structures and technologies on the Web".  Wikipedia +1
- Cybermetrics is a closely-related concept and sometimes used interchangeably, but strictly speaking might refer to quantitative analyses of the broader Internet / cyber-space (not only the Web) (e.g., e-mail, chat, social media).  Karve Institute of Technology +1
- The discipline emerged as an extension of bibliometrics / scientometrics / informetrics, adapted to the Web/Internet context.  IAEME +1
- One of the most visible operational artefacts is the Cybermetrics Lab (at the Spanish National Research Council - CSIC) and its Webometrics Ranking of World Universities.  Wikipedia +2

In short: webometrics seeks to measure the Web (or cyber/internet infrastructure) to infer patterns of production, dissemination, usage, visibility of information (often scholarly) in digital form.

Advantages: What webometrics brings

1. Digital/Online-centric view

- As higher education, scholarly communication, repositories and open access grow online, webometrics gives a lens oriented toward web presence, link structure, domain visibility, repository output, etc., which traditional bibliometrics may miss. (For example, institutional websites, open-access repositories, non-journal outputs).
- The Webometrics ranking emphasises web visibility, openness, and excellence (in its institutional ranking) rather than purely reputation-survey metrics.  UniversityGuru +1

2. Large scale, global coverage

- Because many indicators are derived from publicly-accessible web data (backlinks, domains, repository content, open access materials), one can potentially include many institutions (including those in developing countries) that may not be adequately covered by traditional bibliometric databases. For example the July 2025 Webometrics edition covers $\sim 32,000$ higher education institutions globally.  Figshare +1
- This global reach allows examining digital  es (e.g., how institutions in the Global South fare in web presence) and thus has value for inclusive assessment.  จุฬาลงกรณ์มหาวิทยาลัย... +1

Resultados preliminares: Gemini

- 10 resultados recuperados. Solo 1 resultado (10%) corresponde a un documento académico; el resto son páginas web, rankings y recursos divulgativos.
- 90% de los resultados proceden de sitios web no académicos (portales informativos y páginas sobre Webometrics).
- El único documento científico recuperado contiene términos del prompt en el título, abstract y palabras clave; no está indexado en Scopus ni en Web of Science; corresponde a un preprint publicado en 2025; ha recibido 1 cita en Google Scholar.
- No se recuperan artículos publicados en revistas indexadas ni literatura consolidada sobre Webometrics.
- La mayor parte de la respuesta se construye a partir de fuentes secundarias que describen el ranking Webometrics, en lugar de investigaciones científicas sobre la disciplina.

The discipline of Webometrics (or Cybermetrics) in 2025 has reached a pivotal turning point, characterized by a transition toward greater data transparency and a recalibration of its methodology to align with the Open Science movement.

◆ Cambiar a plan superior

The most significant development is the shift away from proprietary or restricted data sources toward open, sustainable infrastructures.

1. Current State and Methodological Shift

In 2025, the field underwent a "radical" change regarding its primary data sources. Due to access restrictions and the discontinuation of Google Scholar data for automated harvesting, the Webometrics University Ranking migrated the **Openness indicator** from Google Scholar to OpenAlex.

- **Data Infrastructure:** The shift to OpenAlex, utilizing ROR (Research Organization Registry) identifiers, aims to improve accuracy, sustainability, and transparency in bibliometric tracking. [gropedia.com](#)
- **Operational Change:** The traditional Webometrics ranking website ceased public operations in 2025. Data is now distributed via Figshare and other professional channels, often on a paid or formal basis, rather than as a free-access web interface. [www.preprints.org](#)
- **Indicator Refinement:** The methodology has been streamlined into three core pillars:
 - **Visibility (50%):** Measures the number of unique external referring domains (impact of web content). [gropedia.com](#)

Resultados preliminares: Perplexity

- 18 resultados recuperados solo 4 resultados corresponden a documentos académicos y el resto son páginas web, noticias, blogs, rankings institucionales y recursos divulgativos.
- Entre los documentos académicos el 100% contienen términos del prompt en el abstract y las palabras clave y el 75% contienen términos del prompt en las palabras clave.
- 25% están indexados también en Web of Science.
- 100% de los documentos académicos están disponibles en acceso abierto.
- Los trabajos científicos recuperados abarcan el periodo 2008–2024, combinando literatura fundacional con estudios recientes.

 Respuesta  Enlaces

- Examining the latest developments and research trends in Webometrics as of 2025 to provide a compr... >

In 2025, **Webometrics (or Cybermetrics)** remains an active and evolving subfield of *informetrics* that focuses on quantitatively analyzing web-based academic and institutional presence. The discipline continues to play a central role in evaluating global scientific visibility through the **Webometrics Ranking of World Universities**, while also extending its methodologies to broader research domains.

Current State and Focus

Webometrics has expanded its scope, integrating web-based metrics with bibliometric data from platforms such as **Scopus, OpenAlex, and Majestic** to assess around **32,000 higher education institutions** globally in 2025. This multidimensional approach reflects a growing interest in combining *visibility* (web impact) with *academic performance* (publications and citations), reinforcing the field's relevance for global science mapping and digital reputation analysis.  +2

Beyond rankings, current cybermetric studies explore digital governance, web accessibility, and institutional openness. Research lines increasingly intersect with **media and communication studies**, analyzing web impact within public institutions, health informatics, and open-access dissemination.  +1

Resultados preliminares: Elicit

- 10 resultados recuperados (2 falsos o incorrectos)
- 80% contienen términos del prompt en el título
- 100% contienen términos del prompt en el resumen
- La coincidencia en palabras clave muy minoritario
- Promedio de citas en Google Scholar: ~33 citas por documento
- Solo 3 de los 8 documentos académicos identificados están indexados simultáneamente en Scopus y WoS
- La mayoría de los trabajos recuperados son anteriores a 2020
- la media de publicación se sitúa en 2012
- 100% de los trabajos recuperados son de acceso abierto

What is the current state of the Webometrics (also known as Cybermetrics) discipline in 2025, considering its advantages, limitations, current applications, and main lines of research?

Webometrics in 2025 is a multifaceted discipline that uses web-based metrics for research evaluation, institutional ranking, and policy decision-making across various fields, offering advantages like global coverage and interdisciplinary applications while facing challenges including standardization issues, data quality problems, and potential metric manipulation.

Abstract

Webometrics in 2025 appears as a multifaceted discipline defined by both technical innovation and a range of practical applications. Studies report that web-based metrics are used to:

1. Evaluate research and rank institutions by complementing traditional bibliometrics and peer review.
2. Integrate alternative indicators—such as altmetrics and usage metrics—to capture both scholarly influence and broader societal impact.
3. Support policy decision-making and address domain-specific needs in fields such as public health, computing, and communication.

Papers highlight advantages such as novel evaluation methods, broad global coverage, and interdisciplinary applications. They also report challenges including limited standardization, data quality and reliability issues, and the potential for metric manipulation. These findings portray a discipline that is evolving methodologically while remaining subject to issues that affect its robustness as a digital evaluation tool.

Paper search

We performed a semantic search using the query "What is the current state of the Webometrics (also known as Cybermetrics) discipline in 2025, considering its advantages, limitations, current applications, and main lines of research?" across over 138 million academic papers from the Elicit search engine, which includes all of Semantic Scholar and OpenAlex.

We retrieved the 50 papers most relevant to the query.

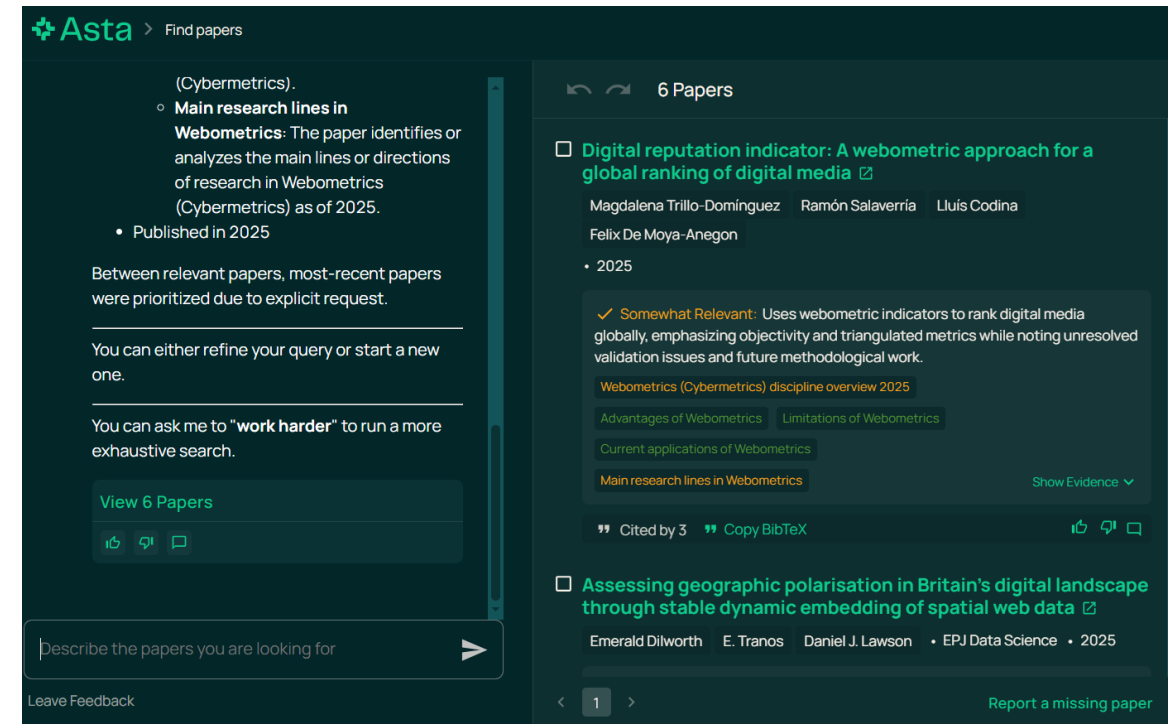
Screening

We screened in sources that met these criteria:

- **Webometrics Focus:** Does this study focus on webometrics, cybermetrics, or web-based bibliometric analysis as a primary topic?
- **Methodological or Applied Content:** Does this study examine methodological approaches, techniques, tools used in webometrics OR describe current applications of webometrics in various domains?
- **Advantages Assessment:** Does this study evaluate, discuss, or identify advantages, benefits, strengths, or positive aspects of webometric methods?
- **Limitations Assessment:** Does this study identify, discuss, or evaluate limitations, challenges, weaknesses, or criticisms of webometric approaches?
- **Future Directions:** Does this study outline emerging trends, future directions, or new research lines in webometrics?

Resultados preliminares: Asta

- 3 documentos recuperados
- 2 de los documentos contienen términos del prompt en el título
- Los 3 contienen términos relacionados con el prompt en el abstract
- Solo 1 de los documentos contienen términos del prompt en las palabras clave
- Solo 1 de los 3 documentos está indexado en Scopus y WoS.
- 67% de los resultados corresponden a preprints o trabajos no indexados en arXive
- Promedio de citas en Scholar :7,3 citas
- Artículo en Wos y Scopus ha recibido 3 citas
- Uno de los artículos no está en acceso abierto



Resultados preliminares: Scopus AI

- 8 documentos recuperados
- 75% de los documentos contienen términos del prompt en el título
- Todos los documentos contienen términos relacionados con el prompt en el abstract
- 62,5% de los documentos contienen términos del prompt en las palabras clave
- Todos los documentos están indexados en Scopus
- 62,5% de los documentos están indexados simultáneamente en Web of Science
- Promedio de citas:
 - 11 citas en Google Scholar
 - 5 citas en Scopus
 - 9 citas en WoS
- Documentos entre 2013 y 2025
- 87,5% de los trabajos son de acceso abierto

What is the current state of the Webometrics (also known as Cybermetrics) discipline in 2025, considering its advantages, limitations, current applications, and main lines of research?

[Show Copilot steps](#)

Summary

Current State of Webometrics in 2025

Advantages:

Impact Analysis: Webometrics is highly effective in analyzing the impact of websites, which is crucial for understanding the reach and influence of academic and institutional web presences 1 2.

Ranking Studies: It is widely used for ranking universities and academic institutions globally, providing a standardized measure of academic performance and visibility 1 2 3.

Interdisciplinary Applications: Webometrics research is cited across various disciplines such as computer science, sociology, physics, and complex science, indicating its broad applicability and relevance 4.

Limitations:

Language Barriers: Non-English content can hinder the impact and visibility of websites, as seen in the case of Iranian universities where Farsi content limited accessibility 1 2.

Theoretical Focus: Much of the webometrics research remains theoretical, which may limit its practical applications unless it demonstrates clear utility 4.

Manipulation Risks: Metrics based on social media attention (e.g. tweets, likes) are prone to manipulation, which can distort the true impact and relevance of the content 5 6.

Current Applications:

University Rankings: Webometrics is extensively used to rank universities, helping institutions understand their global standing and areas needing improvement 1 2 3.

Economic and Financial Analysis: There is emerging evidence of a relationship between webometrics indicators (like inlinks) and economic/financial variables in the corporate sector, suggesting new avenues for management research 7.

Research Evaluation: Webometrics is used alongside traditional scientometric indicators to evaluate research impact, combining scholarly and wider societal impacts 5 6.

Main Lines of Research:

Interdisciplinary Impact: Research is focusing on how webometrics can be integrated with other disciplines to enhance its applicability and relevance 4 8.

Improving Metrics: Efforts are ongoing to refine webometrics indicators to better reflect true impact and reduce susceptibility to manipulation 5 6.

Strategic Frameworks: Developing strategic frameworks to help institutions

Resultados preliminares: WoS AI

- 8 documentos recuperados (todos ellos de 2025/26)
- Ninguno de los documentos contienen términos del prompt en el título, abstract o palabras clave
- El 88% de los documentos están indexados también en Scopus
- 88% de los artículos son de acceso abierto.
- Promedio de citas:
 - 9,4 citas en Google Scholar
 - 4,6 citas en Scopus
 - 1,3 citas en Web of Science
- 100% de las revistas pertenecen a Q1 o Q2

What is the current state of the Webometrics (also known as Cybermetrics) discipline in 2025, considering its advantages, limitations, current applications, and main lines of research?

How are these results generated? [arrow_drop_up](#)

Overview:

The current state of the Webometrics (Cybermetrics) discipline in 2025 reflects a growing interest in its applications, advantages, and limitations, particularly in the context of digital transformation and data-driven decision-making. Key themes include the "integration of advanced technologies", "challenges in data accuracy and reliability", and "emerging research trends and applications".

Integration of advanced technologies:

- The discipline is increasingly leveraging artificial intelligence and machine learning to enhance data analysis and visualization, which improves the understanding of web-based metrics ^[10].
- There is a notable focus on the application of Webometrics in various fields, including healthcare, where it aids in clinical decision support systems ^[18].
- The use of bibliometric analysis tools is becoming common to map research trends and identify influential studies within the Cybermetrics field ^[4].

Challenges in data accuracy and reliability:

- Despite advancements, issues related to the accuracy of data and the reliability of metrics remain significant barriers to the broader adoption of Webometrics ^[10].
- Concerns about the validity of data sources and the methodologies used in Webometric studies highlight the need for standardized assessment frameworks ^[6].
- The integration of diverse data types and the complexity of web environments complicate the establishment of consistent metrics ^[19].

Emerging research trends and applications:

- Recent studies indicate a shift towards exploring the socio-economic impacts of web metrics, particularly in understanding user behavior and engagement ^[23].
- There is an increasing emphasis on interdisciplinary approaches, combining insights from various fields to enhance the relevance and applicability of Webometrics ^[9].
- Future research is expected to focus on developing scalable methodologies that can adapt to the rapidly changing digital landscape ^[17].

Resultados preliminares: Scispace

- 13 documentos recuperados de 2013-2025 con predominio 2024-2025
- 69% de los documentos contienen términos del prompt en el título y en el abstract
- 38% de los documentos contienen términos del prompt en las palabras clave
- 46% de los documentos están indexados en Scopus
- 23% de los documentos están indexados simultáneamente en Web of Science)
- 85% de los documentos son de acceso abierto
- Promedio de citas:
 - 6,3 citas en Google Scholar
 - 5,5 citas en Scopus (entre los artículos indexados)
 - 9 citas en WoS (entre los artículos indexados)

The screenshot displays the Scispace web interface. The main workspace is titled 'Webometrics Discipline Update' and contains a document titled 'Analyzing the current state of Webometrics/ Cybermetrics discipline in 2025'. The document text reads: 'I'll conduct a comprehensive research analysis to provide you with the most current information about Webometrics (Cybermetrics) in 2025, including its advantages, limitations, applications, and research directions.' Below the text, three search queries are listed: 'Searched SciSpace What is the current state of Web...', 'Searched SciSpace Full Text What is the current sta...', and 'Searched Google Scholar webometrics cybermetrics 20...'. A section titled 'Merging and ranking research results for comprehensive analysis' shows a query: 'Combined and reranked results current state of Web...'. On the right sidebar, there are tabs for 'Live Activity' and 'All Files'. Under 'All Files', a folder named 'Final Outputs' is expanded, showing a file named 'webometrics_analysis_20...'. At the bottom of the interface, there is a promotional banner for 'New to SciSpace? Sign up and get 100 Credits for free' with a 'Sign up' button. A 'Download All Files' button is also visible at the bottom right.

Resultados preliminares: Epsilon

Search

You are not signed in.

🔗 Leave Feedback

👤 Log In / Sign Up



What is the current state of the Webometrics discipline in 2025, including its advantages, limitatio...

From The Literature Below

📄 Copy 🔄 Share ☆ Save

- **Current State and Applications:** In 2025, webometrics is significantly enhanced by artificial intelligence (AI), which improves web crawling, data collection, web link analysis, web content analysis, social media analysis, web impact analysis, and recommender systems [2.2](#) . AI integration allows for more objective and reliable analyses in webometrics, similar to its impact on scientometrics and bibliometrics [2.2](#) .
- **Advantages:** The integration of AI in webometrics offers enhanced data collection and analysis capabilities, leading to more accurate and efficient insights into web-based scholarly communication [2.2](#) . AI also facilitates automation in data collection and analysis, which is a significant advantage in handling large datasets [2.2](#) .
- **Limitations:** Despite its advancements, webometrics faces limitations such as the reliance on internet access and the use of specific search engines, which may not reflect the entire population's behavior [8.6](#) . Additionally, the field must address challenges related to data transparency and replicability [8.5](#) .
- **Main Lines of Research:** Current research in webometrics focuses on optimizing AI algorithms for better data analysis and exploring new applications in web-based scholarly communication [2.1](#) . There is also a growing interest in integrating webometrics with other metric studies like infodemiology to predict and analyze public behaviors and trends [8.1](#) .

▼ References

- [2](#) 📄 Unleashing the Power of AI. A Systematic Review of Cutting-Edge Techniques in AI-Enhanced Scientometrics, Webometrics, and Bibliometrics
- [8](#) 📄 OPEN ACCESS EDITED BY

- 2 documentos recuperados ambos en acceso abierto
- 100% de los documentos contienen términos del prompt en el título y el abstract
- 50% de los documentos contienen términos del prompt en las palabras clave
- Indexados en Scopus y Web of Science (Q1,Q2)
- Los trabajos recuperados son muy recientes (2023–2025)
- Promedio de citas:
 - 70 citas en Google Scholar
 - 33 citas en Scopus
 - 35 citas en WoS.
- Las revistas identificadas pertenecen a áreas directamente relacionadas con la información científica y los datos

Principios del ALO

Optimización semántica

- Optimización semántica
- Títulos explícitos y descriptivos
- Abstracts ricos en conceptos
- Alta coincidencia entre título, resumen y palabras clave
- Terminología consistente en todo el artículo

Optimización bibliográfica

- Publicar en revistas indexadas (Scopus/WoS)
- Favorecer revistas Q1/Q2 cuando sea posible
- Artículos fundacionales del campo

Optimización digital

- Acceso abierto
- Presencia en repositorios
- Google Scholar

Patrones que aparecen en las nueve herramientas

El abstract es el rey

- la coincidencia en el abstract es superior a la del título y muy superior a las keywords
- El abstract parece ser el principal punto de entrada para los LLM

Las keywords importan... menos de lo esperado

- Las coincidencias en keywords son mucho menores
- Esto sugiere que los LLM trabajan sobre la representación semántica del texto completo y no sobre los metadatos clásicos.

Patrones que aparecen en las nueve herramientas

El acceso abierto aparece constantemente

- Prácticamente todas las plataformas recuperan mayoritariamente artículos Open Access.

Los papers fundacionales aparecen repetidamente

- Aguillo y Thelwall como referencias clásica de Webometrics

Patrones que aparecen en las nueve herramientas

A partir de los resultados preliminares de las **nueve herramientas** sugerimos la siguiente **hipótesis**:

“La visibilidad de un artículo científico en los modelos de lenguaje depende de la combinación de tres dimensiones : optimización semántica, autoridad científica y presencia en el ecosistema digital”

Limitaciones

- Más herramientas de IA: Claude, Consensus, Google Labs
- Diferentes modelos de IA
- Herramientas en versión freemium
- La constante evolución de esta tecnología



Muchas gracias por vuestra atención: Estoy abierto a cualquier pregunta y/o sugerencia

Dr. Carlos Lopezosa

PrimeTime

Jornada de transferencia de los resultados del proyecto

Valencia, 26 de junio 2026



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA