Scopus AI: Trusted content. Powered by responsible AI.

June 25, 2024

Annapaola Migani, PhD

Contents

1. Introducing Scopus AI
2. Why Scopus AI?
3. How does it work?
4. Scopus AI Demo
5. User Feedback
6. Scopus foundations

Introducing Scopus AI

Your dynamic GenAI-powered research companion

There are many challenges in research today

Early career researchers and academics

Complex research landscape	Limited search tools	Disciplinary silos
Information overload	Inefficient search	Limited collaboration

Suboptimal productivity and collaboration / missed opportunities

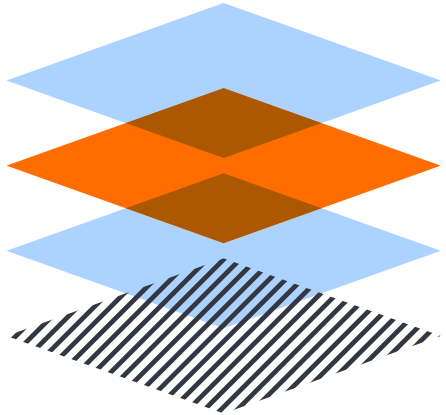
Lower academic and societal impact

Current tools fail to get you up to speed quickly

Current tools fail to give a focused, easy-to-use summary of unfamiliar research fields, making it difficult to quickly understand important ideas and find connections with potential collaborators.

While traditional tools falter, Scopus AI with generative AI aims to offer brief and reliable summaries of research topics.

This will help you conduct research more efficiently, find relevant information, and facilitate effective collaborations across different disciplines.



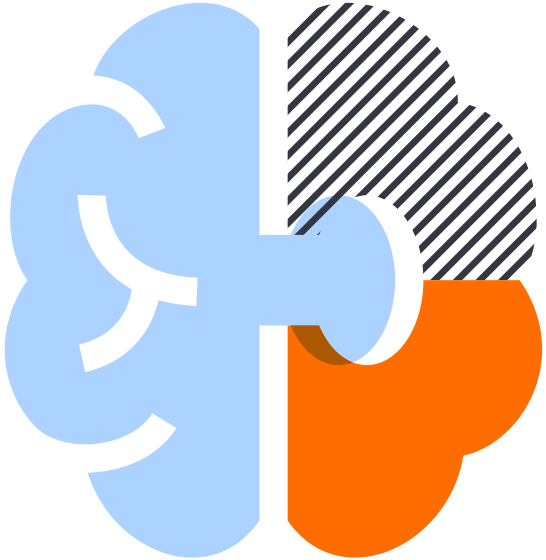
Introducing Scopus AI

Meet Scopus AI.

Our new AI-powered academic search tool uses top-notch peer-reviewed research to provide you with focused and easy-to-understand insights on new terms, concepts and disciplines.

We are collaborating with the research community to lead the development of innovative generative AI techniques for researcher tools.

Join us! Let’s discover this together.



What use cases does Scopus AI support?

- **Finding references**
  (e.g., for thesis or manuscript)
- **Getting familiar** with a new topic
- **Deepening understanding** of a research topic
- **Better articulating** a hypotheses or ideas
- **Explaining a complex topic**
  to juniors / non-academics
- **Collaborating** with others

Scopus AI at a glance

Referenced Summary

Concept map

Foundational papers

Topic experts

Natural language queries

Expanded referenced summary

Go deeper questions

* Interface and functionality subject to change

“Something that gives me a lot of confidence in [Scopus AI] is that the **results are referenced**, and it’s easy to check the references.”

Bruno Augusto
Environmental Engineering PhD student**,**
University of Aveiro, Portugal

Read the full interview

Why Scopus AI?

We explore some of the features and principles that make Scopus AI so unique

Scopus AI champions **responsible AI**

In 2022, we published 5 Responsible AI Principles that complement our existing AI policies and processes.

These principles have shaped every stage of Scopus AI's development, and will continue to guide the tool's evolution.

Our Responsible AI Principles:

1. We consider the real-world impact of our solutions on people.
2. We take action to prevent the creation or reinforcement of unfair bias.
3. We can explain how our solutions work.
4. We create accountability through human oversight.
5. We respect privacy and champion robust data governance.

https://www.elsevier.com/about/policies-and-standards/responsible-ai-principles

Scopus AI uses only **high-quality, curated Scopus content**

Every Scopus AI response is grounded in content that has been vetted by independent experts.

It has not only been peer reviewed, it has also been rigorously vetted and selected for inclusion in Scopus by the independent Content Selection and Advisory Board (CSAB).

Journals must demonstrate their ability to maintain their quality status every year as part of the Scopus reevaluation program.

Stage 1	**~3,500** title suggestions per year on average
Stage 2	**~51%** meet the Scopus minimum criteria
Stage 3	**~48%** are accepted after the CSAB's review
Result	**~857** serial titles meet the full Scopus criteria

Scopus AI takes extensive steps to **minimize hallucinations**

The large language models (LLMs) used in GenAI tools bring many benefits, but they also come with shortcomings.

These include the potential to generate '**hallucinations**' — inaccurate or false responses, undermining trust in the information they deliver.

Scopus AI takes a multi-stranded approach to reducing hallucinations.

1. **Responses are grounded in Scopus content**
2. **Sources are updated daily**
3. **Strict 'guardrails' guide the LLM**
4. **Sophisticated retrieval augmented generation (RAG) fusion technology**
5. **Rigorous evaluation frameworks**
6. **Cross-departmental human oversight**
7. **Community feedback**

“There’s a lot of poor information that gets generated by other, less specialized tools. Even though they’re written in a way that seems true, the results are often very inaccurate or outright false.

Using Scopus AI, you get information that is generated from abstracts that come from **peer-reviewed journals**. So mostly you get trusted information.”

Read the full interview

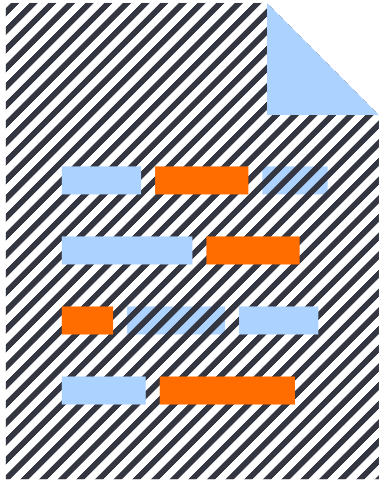
Dr Engie El Sawaf
Pharmacology Lecturer Assistant,
Future University, Egypt

Scopus AI is **transparent** and always shows its workings

We've made Scopus AI responses easy to evaluate, verify and confirm.

Every claim or assumption cites the Scopus documents used to generate it. And because Scopus AI is built on Scopus, the references can be downloaded in bulk to your preferred reference manager, or to SciVal for further analysis.

Scopus AI provides guidance on its confidence in the relevancy of its response and can explain the technology it uses**.**



Scopus AI takes action to **prevent unfair bias**

Scopus AI is committed to preventing the creation or reinforcement of all forms of unfair bias.

We draw on a potent mix of guidelines, evaluation, technology and human oversight to help us achieve this.

For example, we actively test Scopus AI for biased responses using a variety of internationally-recognized frameworks.

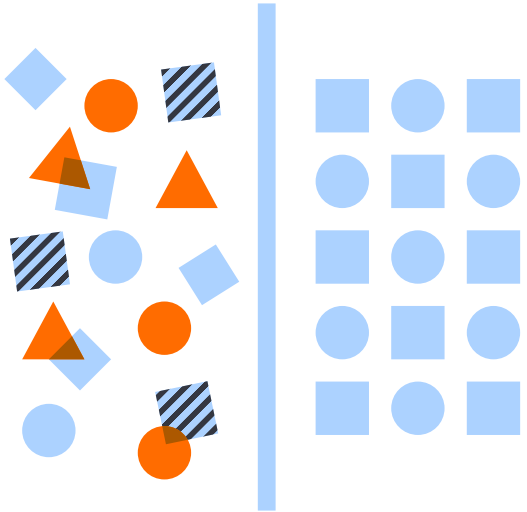
1. **We use only trusted Scopus data**
2. **We focus on relevancy of content, not impact**
3. **We conduct rigorous evaluations**
4. **We provide the LLM with strict guidance**
5. **We learn from user testing**

Scopus AI **respects data privacy**

We have designed Scopus AI to avoid unnecessary data retention.

For example, Scopus AI uses OpenAI's large language model (LLM) ChatGPT hosted on Microsoft Azure to synthesize results found by our vector search. We have an agreement in place that information will not be stored or used to train OpenAI's public model.

We also comply with privacy regulations like GDPR.

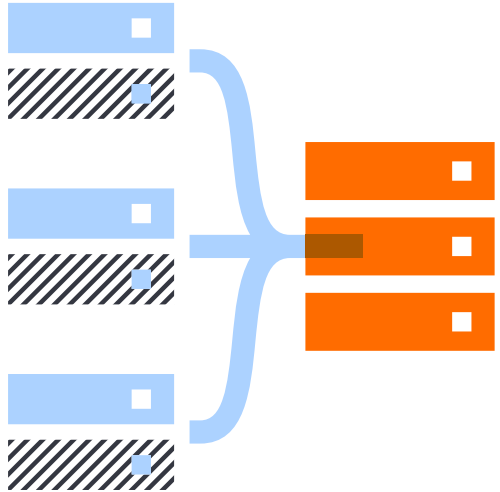


Scopus AI draws on a **unique & powerful blend of technology**

Scopus AI's strength lies in the way these technological elements interact with the underlying dataset.

For example, the sophisticated **vector search** and custom-built **small language model** source the most relevant Scopus content**,** while the **prompt engineering** ensures the quality of the large language model's response.

In addition, **Scopus AI is home to our patent-pending retrieval augmented generation (RAG) fusion technology**.



Scopus AI was developed **with the academic community**

Decisions around Scopus AI enhancements are rooted in compelling user feedback.

We have worked with thousands of researchers, academic leaders and librarians worldwide to understand their needs and rigorously test new ideas since the earliest design stages of Scopus AI.

That collaboration remains just as valuable today, as we continue to evolve Scopus AI.

Just some of the features shaped by user feedback

- Position of Scopus AI in Scopus
- Emphasis on references
- Concept map
- Foundational papers
- Expanded summary
- Confidence layer
- SciVal export option

How do we ensure quality and reliability?

Scopus content

Trusted, high quality content

The content that Scopus AI draws on is peer reviewed and has been rigorously vetted and selected for inclusion in Scopus by the independent Content Selection and Advisory Board (CSAB).

Re-evaluation policy

Selected journals must also demonstrate the ability to maintain their quality status year over year as part of the Re-evaluation program.

Scopus AI platform

RAG fusion

Retrieval augmented generation (RAG) fusion is an approach that improves the quality of both vector search retrieval and the generation of LLM summaries. Together, these factors reduce the risk of hallucinations (or false AI-generated information).

Prompt engineering

This provides the large language models with strict instructions about which content they should use to generate their response to your query. Limits the risk of hallucinations — or false AI-generated information — ensuring reliable answers.

Reflection layer

Reflection layer ensures appropriate content and nuance in summary outputs. Our prompting aims to inform the user of the confidence of our responses based on the content provided.

Scopus AI output

Quality evaluation framework

This checks responses for relevancy, coherency and safety/reliability.

Harmful evaluation framework

Regular testing keeps our prompts engineered to avoid bias and unreliable results.

Adheres to responsible AI principles & data privacy policies

Developed responsibly

For over 10 years, Elsevier has used AI and machine learning responsibly in our products, combining it with unparalleled peer-reviewed content, extensive data sets and sophisticated analytics.

Scopus AI is developed in line with Elsevier's Five Responsible AI Principles.

Data privacy

Scopus AI adheres to Elsevier's Privacy Principles and GDPR regulations to guarantee user privacy.

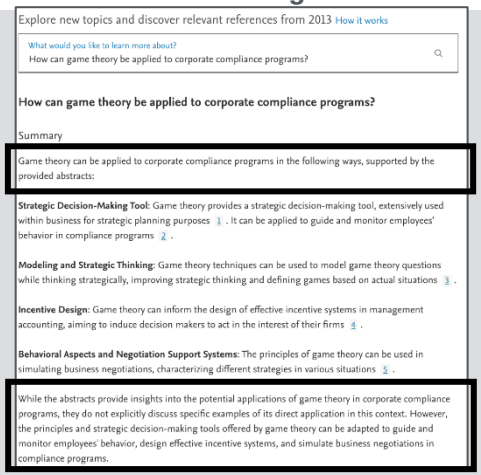
Our use of OpenAI's LLM is private so there is no data exchange.

The product has also been designed to avoid unnecessary data retention.

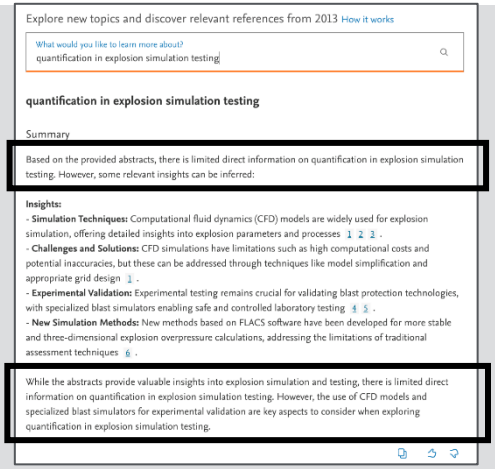
Transparency and context about our confidence in the responses

Reflection layer: Ensures appropriate context and nuance in summary outputs. Our prompting informs the user of the confidence of our responses based on the content provided.

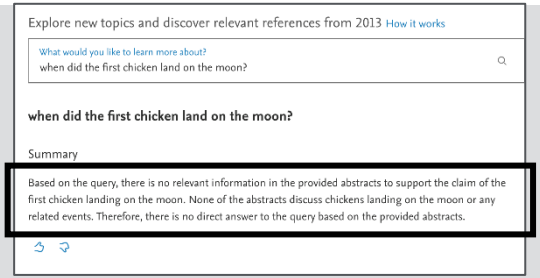
Direct responses are made when the **confidence** in the source material is **high**



Inferred responses are made when **confidence** in source material is **medium**



No response is provided when **confidence** in source material is **low**



Nuance is added to provide transparency to users about the nature of available information

*Interface and functionality subject to change

Addressing researcher discovery needs: Gen AI tools

“Generalist” Gen AI tools

Value:
- ✓ Sophisticated LLMs
- ✓ Free versions

Watch outs:
- Answers can be given from generic, non-peer reviewed content.
- Can create references that do not exist
- Outside research workflow
- Combining with academic literature provides mixed results, gaps, extraneous content
- Trained on older data
- Privacy

Scopus AI

Value:
- ✓ Available in Scopus, seamlessly within the research workflow
- ✓ Uses reliable curated peer-reviewed content
- ✓ Leverages patent pending technology
- ✓ Leverages cutting edge tech from Chat GPT with guardrails specially built for the academic community’s information and decision-making needs
- ✓ Developed using responsible AI principles for transparency and trust
- ✓ Via evaluation & quality frameworks, minimal hallucinations

Other “Academic” Gen AI tools

Value:
- ✓ Free versions exist
- ✓ Some citable sources

Watch outs:
- Content can be limited or non-transparent
- Can include duplicate data
- Can include less trusted content
- May be outside current research workflow tools
- Require additional login/access

Trusted content
…that champions academic integrity

Cutting-edge technology
…that saves you time

Focused results
…for reliable, focused insights

How does it work?

Our 6-step process

How does it work?

Step 1:
Curation of high-quality Scopus content

Step 2:
Query formulation

Step 3:
Scopus AI: Vector Search & Result Generation

Step 4:
Large Learning Model (LLM) summary generation

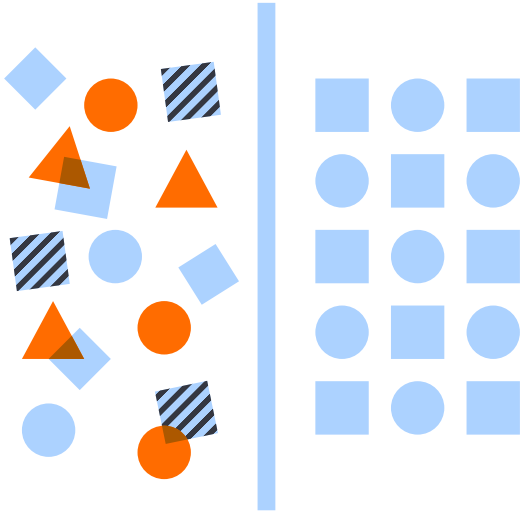
Developed according to our **Responsible AI Principles**[1]

Explore these **4 steps** in greater detail on the following slides.

Curation of high-quality Scopus content

The Scopus collection of high-quality, reliable academic content informs the results

1. Scopus AI is informed by **the high-quality, reliable, curated academic content** that the research community knows and trusts.
2. Data is **updated daily**, ensuring the most recent content is included.
3. Abstracts of **all content types** are used for summaries, and prompt engineering minimizes the use of irrelevant documents.

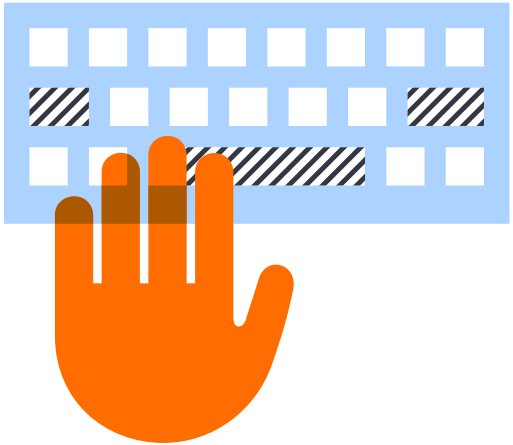


You can be confident that your summary is based on expert-vetted, curated information, generated via a transparent content selection process

Query formulation

You can alter your query type depending on the nature of your question and the level of information you have on the topic

- **You define the query**!
- You can add a simple keyword or phrase or a more complex paragraph, title or part of an abstract.



With no search string requirements, you can search based on any information you have

“Before you even start building your knowledge in an area, you have to know a lot of technical terms. But with Scopus AI, **you can use more informal language when searching**, and you get pretty much the same results as you do with the technical terms. I think that’s one of the great benefits of using the tool.”

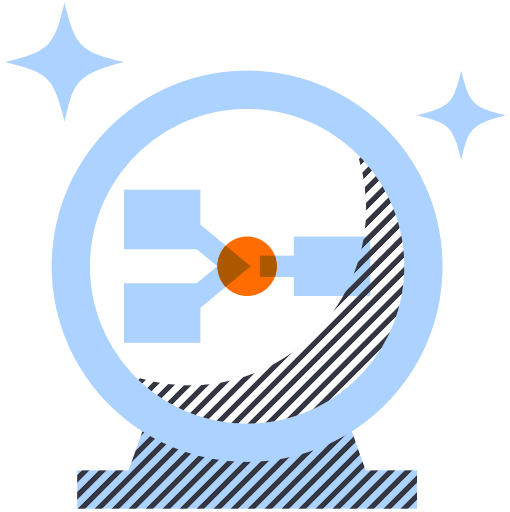
Read the full interview

Bruno Augusto
Environmental Engineering PhD student**,**
University of Aveiro, Portugal

Scopus AI: Vector Search & Result Generation

When you submit a query, Scopus AI immediately gets to work

- Vector transformation: Your query is swiftly transformed into a vector using our **advanced language model**.
- Vector search: We pinpoint the most **pertinent academic documents** based solely on their abstracts.
- Relevance prioritized: Our proprietary **semantic relevancy algorithm** accounts for several factors, including **cosine similarity** (a trusted indicator of relevance), **the recency of the publication, and citation count** to ensure **the quality and relevance** of the results.



Our sophisticated approach ensures you receive only the most relevant and authoritative responses

Step 3: Scopus AI generates your summary

The results of the vector search are fed into our large language model (LLM), along with your original query, and our engineered prompt.

The LLM is tasked with taking the information that has been shared with it and synthesizing it to create the Summaries.

Prompt engineering provides the LLM with clear rules it must follow while generating the response.

Results are regularly checked for quality and safety using sophisticated frameworks.

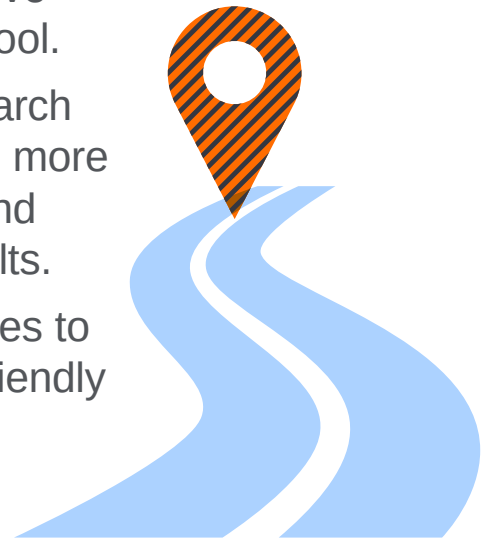
How is Scopus AI evolving?

Since Scopus AI's launch in January 2024, we've drawn on community feedback to introduce:

- Guidance on Scopus AI's confidence in its response.
- The option to export references to SciVal.
- A new small language model reranker that significantly enhances the precision of our search capabilities, ensuring that the most relevant Scopus content surfaces first.

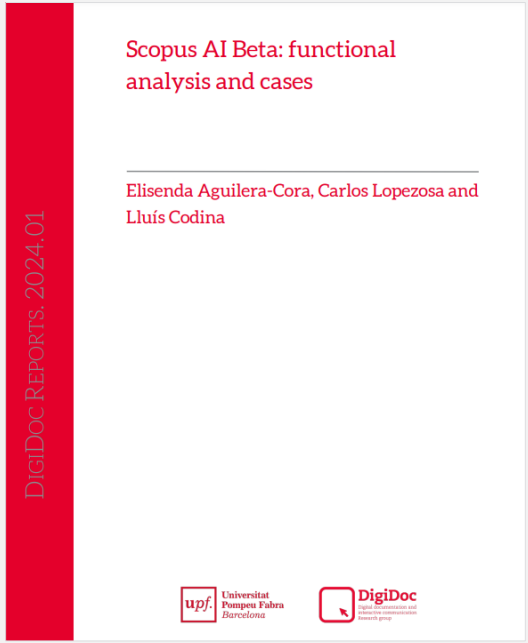
Future developments in the pipeline include:

- Enhancements to query functionality supporting more intuitive and interactive interaction with the tool.
- Improvements to search algorithms to ensure more accurate, relevant and comprehensive results.
- User interface updates to make it more user-friendly and accessible.



Scopus AI Demo

What are users saying about Scopus AI?



Read the full preprint

“ The Scopus AI interface is intuitive and easy to use, it allows the researcher to obtain an overview of a problem, as well as identify authors and approaches, in a more agile search session than conventional search.”

“ It is a valuable tool for literature reviews, construction of theoretical frameworks and verification of relationships between variables, among other applications that are actually impossible to delimit.”

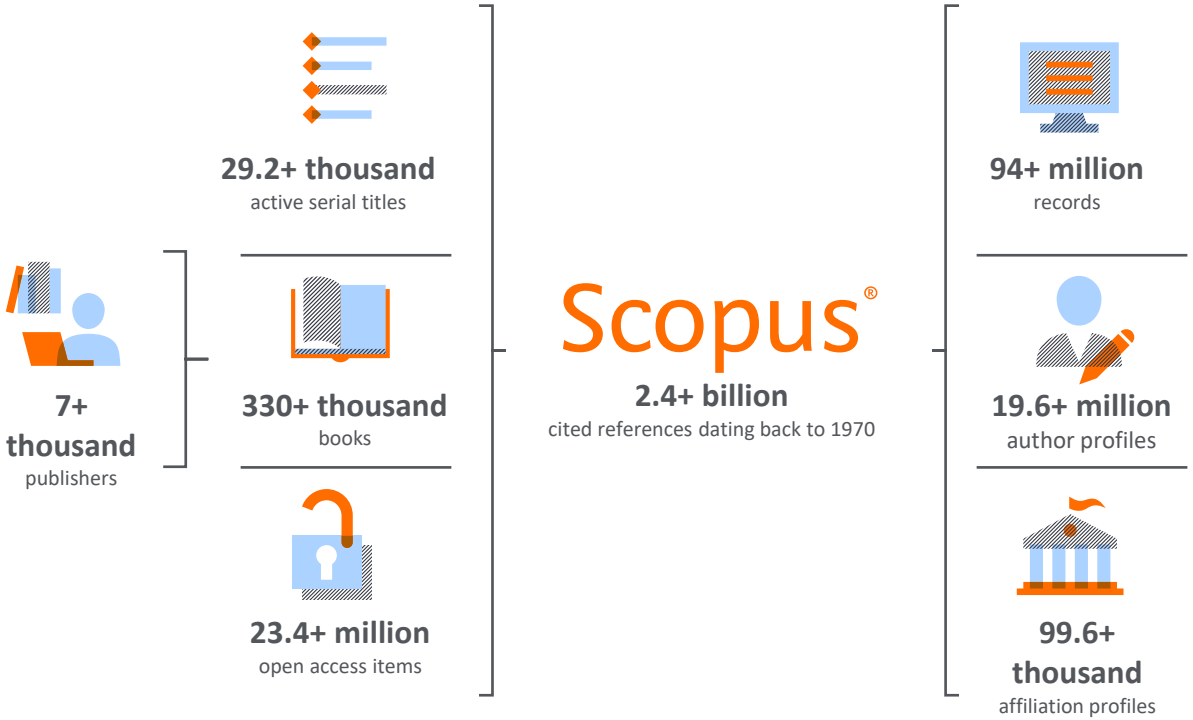
Elisenda Aguilera
Researcher at Pompeu Fabra University in Spain

Scopus foundations

Curated, enriched and connected data that surfaces signals about research that are **intuitive to access and understand**

ELSEVIER

Enhance research and scholarship with comprehensive data and analytics



ELSEVIER

Comprehensive coverage

Globally sourced

- **7,000+ publishers**
- **105 countries**
- **40 languages**

Format and historically inclusive

- **More curated, global content** than competitors
- Historical coverage **back to 1788**
- **23.4+ M open access documents**
- Multiple regional content types **(journals, conferences, books, book series)**

Current

- **Updated daily**

Global Representation

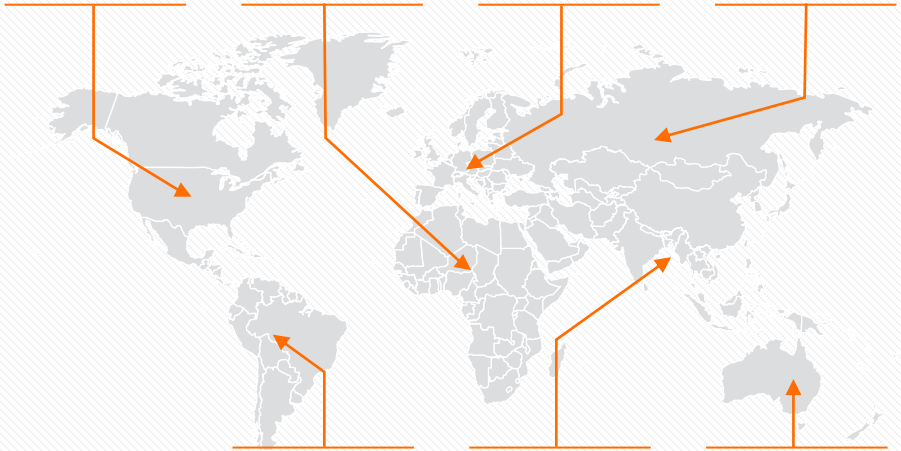
(number of active titles)

North America
6,968
128% more than nearest competitor

Middle East+Africa
1,017
184% more than nearest competitor

Western Europe
13,819
141% more than nearest competitor

East Europe incl. Russia
2,282
232% more than nearest competitor



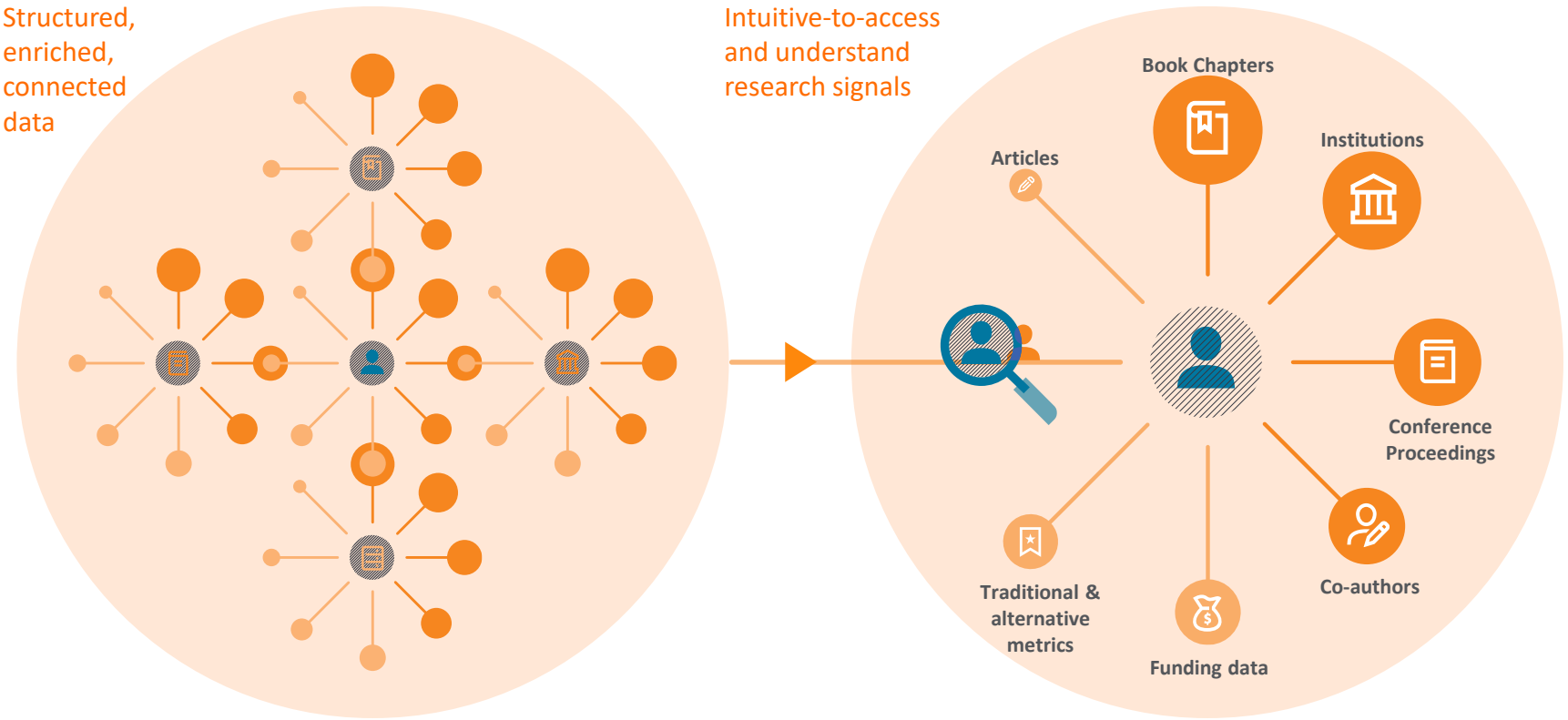
Latin America
994
210% more than nearest competitor

Asia Pacific
2,687
246% more than nearest competitor

Australia/ New Zealand
231
171% more than nearest competitor

Coverage Guide

Curated, enriched and connected data delivers insights about research that are confidence-inspiring, informed by context and intuitive to access



Vetted by independent experts

Scopus Content Selection and Advisory Board (CSAB)

- Independent board of subject experts from all over the world
- Comprised of 17 Subject Chairs
- Chosen for their expertise in specific subject areas; many have (journal) Editor experience.

Selection and reevaluation process

- Rigorous and transparent quality and ethics selection criteria used to evaluate potential titles
- Regularly revaluates Scopus content and discontinues titles no longer meeting the guidelines

Stage 1

~3500

title suggestions per year on average

Stage 2

~51%

meet the Scopus minimum criteria

Stage 3

~48%

are accepted after the CSAB's review

~857

Serial titles meet the full Scopus criteria

ELSEVIER

Coverage Guide

Scopus Coverage Summary

Global representation means global discovery across all subjects and content types

94.4M records from **29.2K** serials, **152K** conferences and **334K** books

from more than **7,000** publishers in **105** countries

- Updated daily—**13,000+** new articles per day indexed
- **23.4M+** open access documents
- “Articles in Press” from **>8,740** titles
- **2.0M+** preprints from multiple preprint servers
- **7,115** active Gold Open Access journals indexed

Number of journals by subject area**

Physical sciences
15,151

Health sciences
14,940

Social sciences and humanities
15,446

Life sciences
8,094

Journals

28,153* active peer-reviewed journals

7,115 Gold OA Journals (DOAJ/ROAD)

21.1M fully-indexed funding acknowledgements

2.03M preprints

- Full metadata, abstracts and cited references (refs post-1970 only)
- Citations back to 1970

Conferences

158K conference events

12.1M+ conference papers

12.9% of database items

Mainly Engineering and Computer Sciences

Books

74.3K individual book series volumes

330K+ stand-alone books

2.89M total book items

Focus on Social Sciences and A&H

Patents

50.5M patents

5 major patent offices:

- WIPO
- EPO
- USPTO
- JPO
- UK IPO

*Journals may be classified in multiple subject areas: this count includes current actively indexed titles only
**These counts include both active and inactive titles; total number of Scopus journals in database including inactive titles is 43,400

Updated December 2023

Thank you