ML meets Materials

A Bibliometric Deep Dive

Ronaldo Cristiano Prati, Universidade
Federal do ABC,
ronaldo.prati@ufabc.edu.br

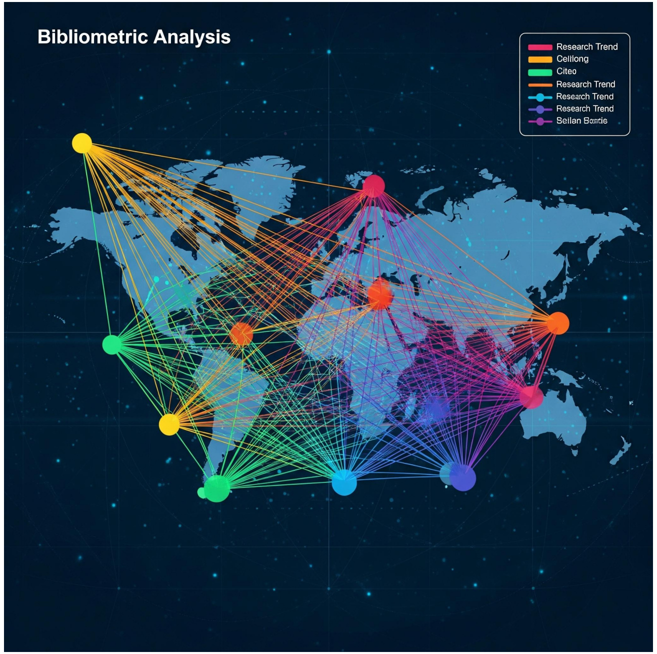
The Importance of Machine Learning in Materials Science

- Accelerates material discovery and design.
- Reduces experimental costs and time.
- Predicts material properties with high accuracy.
- Enables inverse design for targeted applications.
- Uncovers complex structure-property relationships.

The 4th Paradigm

- Data-intensive science is the fourth paradigm
- It involves exploration through massive datasets
- The paradigm integrates theory, experiment, and simulation
- Scientific discovery is increasingly driven by data

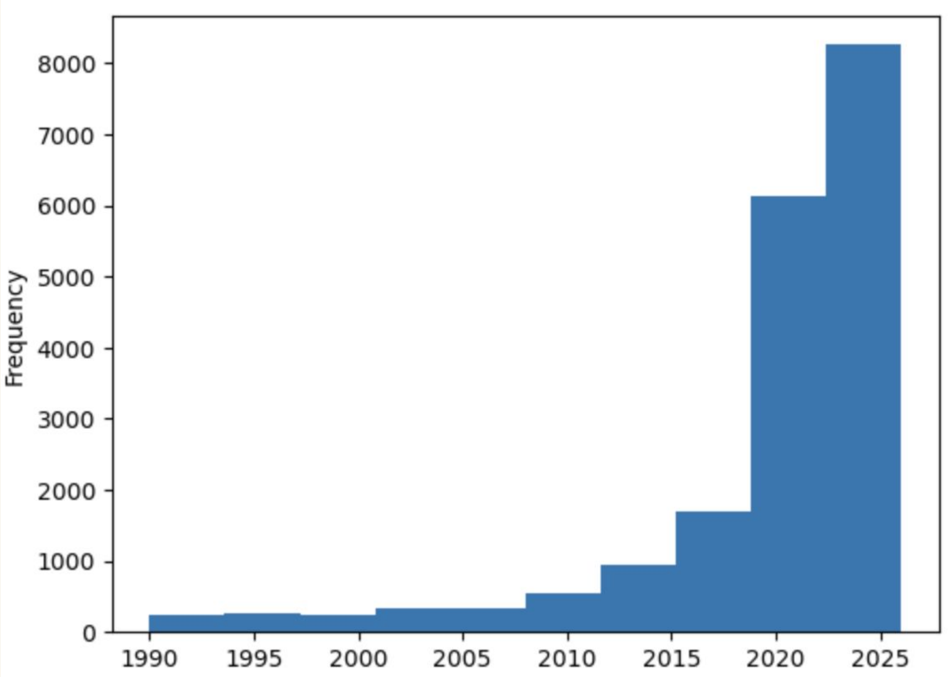
Bibliometric Analysis Essentials



- Bibliometric analysis helps identify important research trends
- It reveals the geopolitics of scientific production globally
- The analysis assists in mapping collaboration networks effectively
- It is crucial for understanding the landscape of science

Exponential Growth of Publications

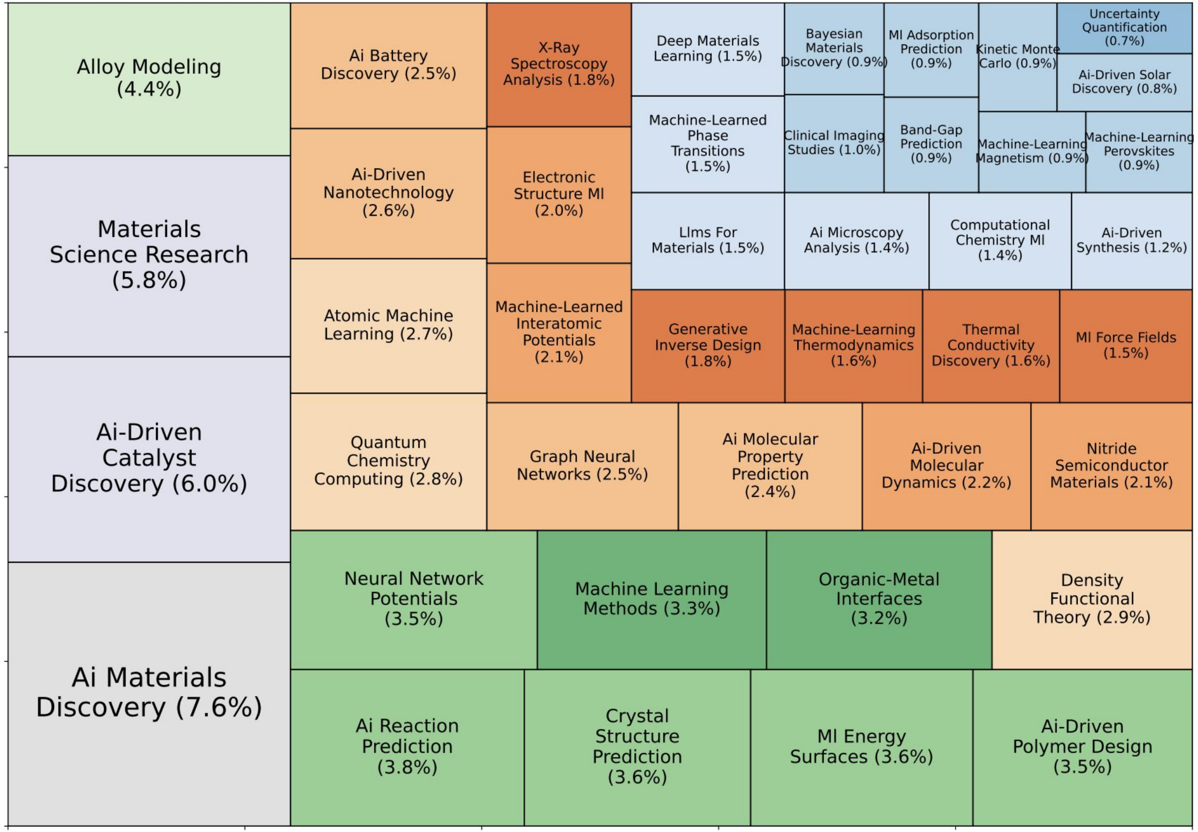
- Dramatic increase in research output in recent years.
- Reflects growing interest and adoption of ML techniques.
- Driven by data availability and computational power.



Data Source and Methodology

- Data source: OpenAlex database.
- Scope: Papers categorized under Machine Learning in Materials Science.
- Focus: Journal papers with an Impact Factor.
- Timeframe: Published after 1990.
- Total volume: 20,127 papers analyzed.

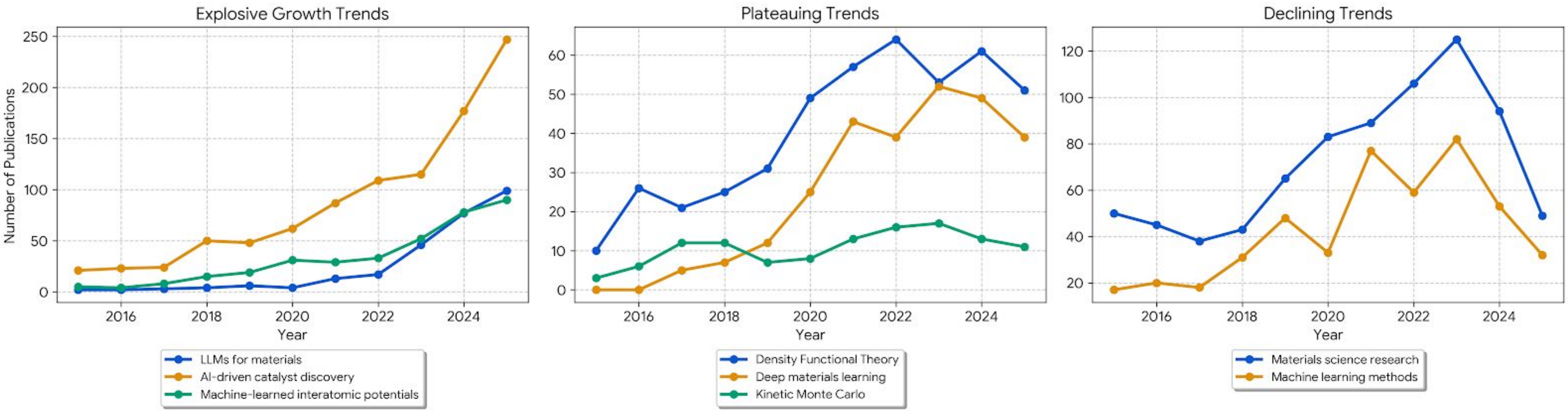
Topic distribution



Key Trends: AI in Materials Science Research

- **Dominance of 'Ai Materials Discovery'**: The largest research category (7.6%), emphasizing the search for novel materials.
- **Foundational Role of ML Methods:** 'Graph Neural Networks' and general 'Machine Learning Methods' are key enabling techniques.
- **Focus on High-Impact Applications**: Significant research in 'Catalyst Discovery', 'Battery Discovery', and 'Polymer Design'.
- **Widespread Adoption**: AI is integrated across diverse fields, from 'Alloy Modeling' to 'Quantum Chemistry'.

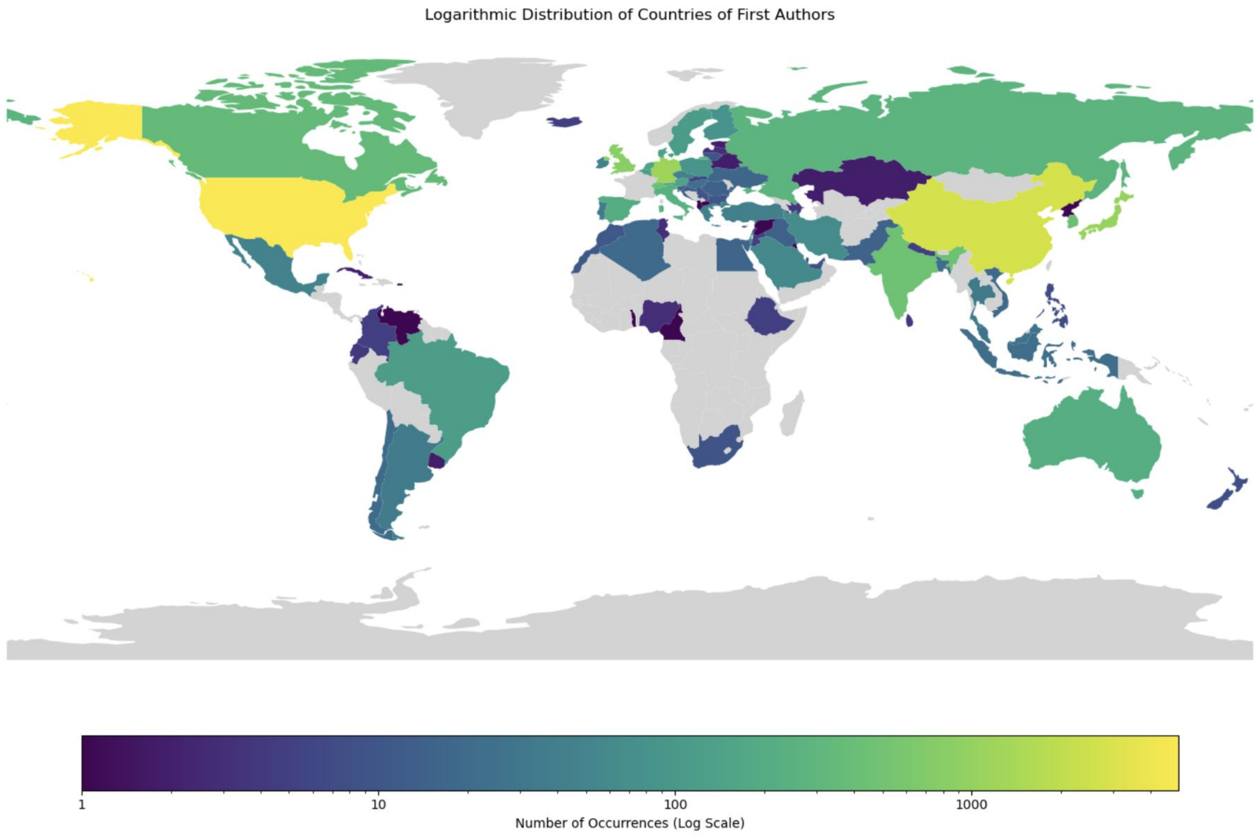
Interesting trends



Highlights: Trends in Materials Science Research Topics

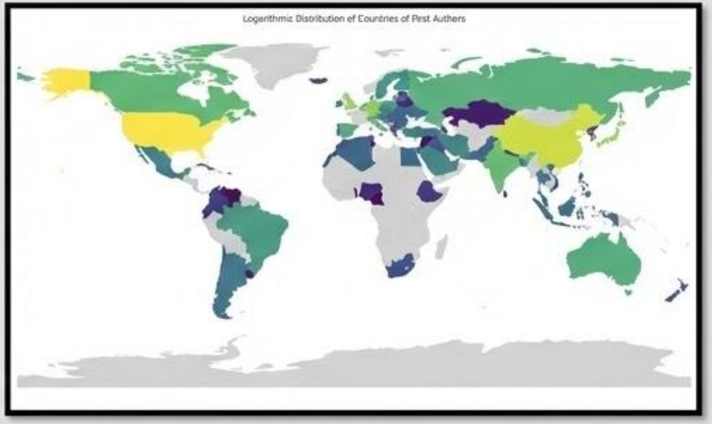
- 🚀 **Explosive Growth in AI-Driven Fields:** Topics like “LLMs for materials,” “AI-driven catalyst discovery,” and “Machine-learned interatomic potentials” are experiencing a rapid surge in publications, particularly since 2022.
- 〜 **Plateauing of Established Methods:** Traditional computational methods such as “Density Functional Theory” and “Kinetic Monte Carlo,” along with earlier “Deep materials learning” approaches, appear to have reached a **saturation point** in research output.
- ↘ **Shift Towards Specialization:** Broader categories like general “Materials science research” and “Machine learning methods” are seeing **a decline**, suggesting a research focus shift towards more **specific**, applied AI and ML techniques in materials science.

Distribution of first authorship

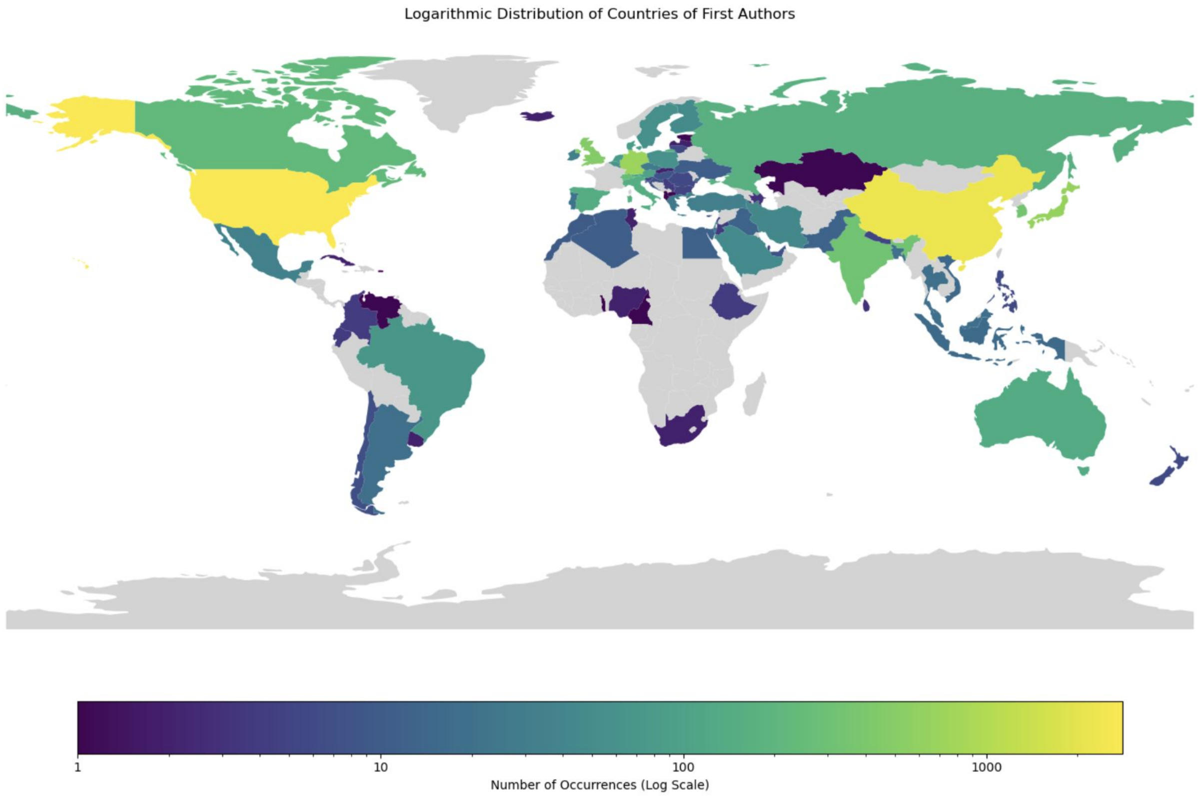


Analysis of Global First Author Distribution

- **Dominant Research Hubs:** The USA, China, and Western Europe (e.g., Germany, UK) exhibit the highest concentrations of first authors, reflecting their leading global research positions.
- **Key Regional Contributors**: Significant contributions are also evident from nations like India, Japan, South Korea, Brazil, and Australia, indicating strong regional research activity.
- **Global Disparities**: A marked inequality in research output is observable, with lower participation rates across much of Africa, South America, and Central Asia.
- **Logarithmic Scale Effect**: The logarithmic coloring highlights the massive order-of-magnitude differences between top-tier publishing countries and others.



Distribution of authorship (last 5 years)



Comparative Analysis: Authorship Distribution (Since 1990 vs. Last 5 Years)

- **Since 1990 (Long-Term Trend):**
  - Historical dominance by traditional research hubs (USA, Western Europe).
  - Gradual growth in contribution from emerging economies.
- **Last 5 Years (Recent Trend):**
  - Rapid acceleration of China's authorship, establishing it as a top-tier contributor.
  - Increased prominence of countries like India, Brazil, and South Korea.
  - Persistent global disparities remain, with limited change in many developing regions.
- **Key Shift:**
  - A discernible shift in the center of gravity of scientific production towards Asia.
  - High-volume output remains concentrated in a relatively small number of nations, despite broader global participation.

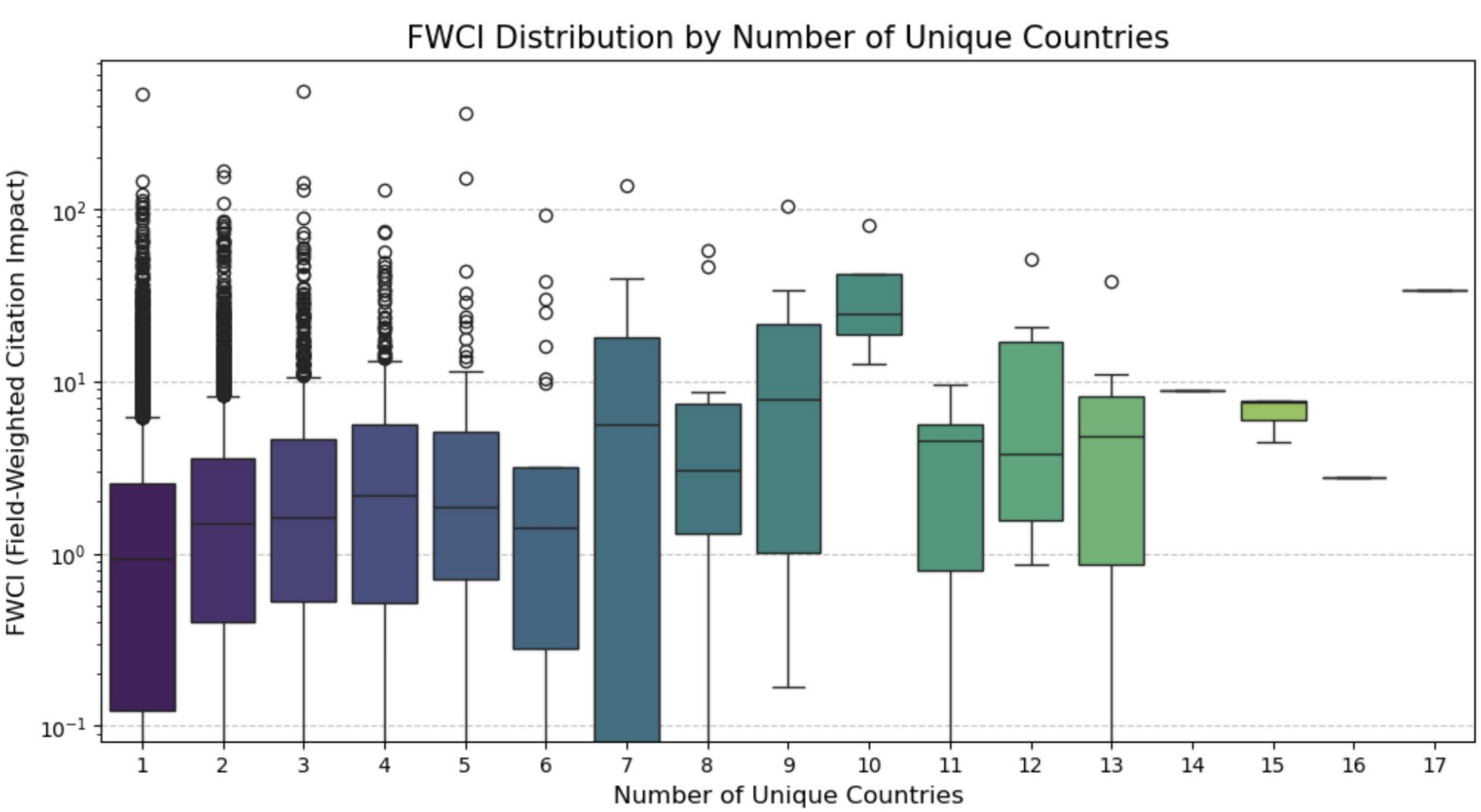
Key Bibliometric Indicators: FWCI & SJR

Field-Weighted Citation Impact (FWCI)

- Measures citation impact compared to world average (1.0) for the same field, publication year, and document type.
- A value > 1.0 indicates higher than average impact; < 1.0 indicates lower than average.
- Accounts for variations in citation practices across different disciplines.

SCImago Journal Rank (SJR)

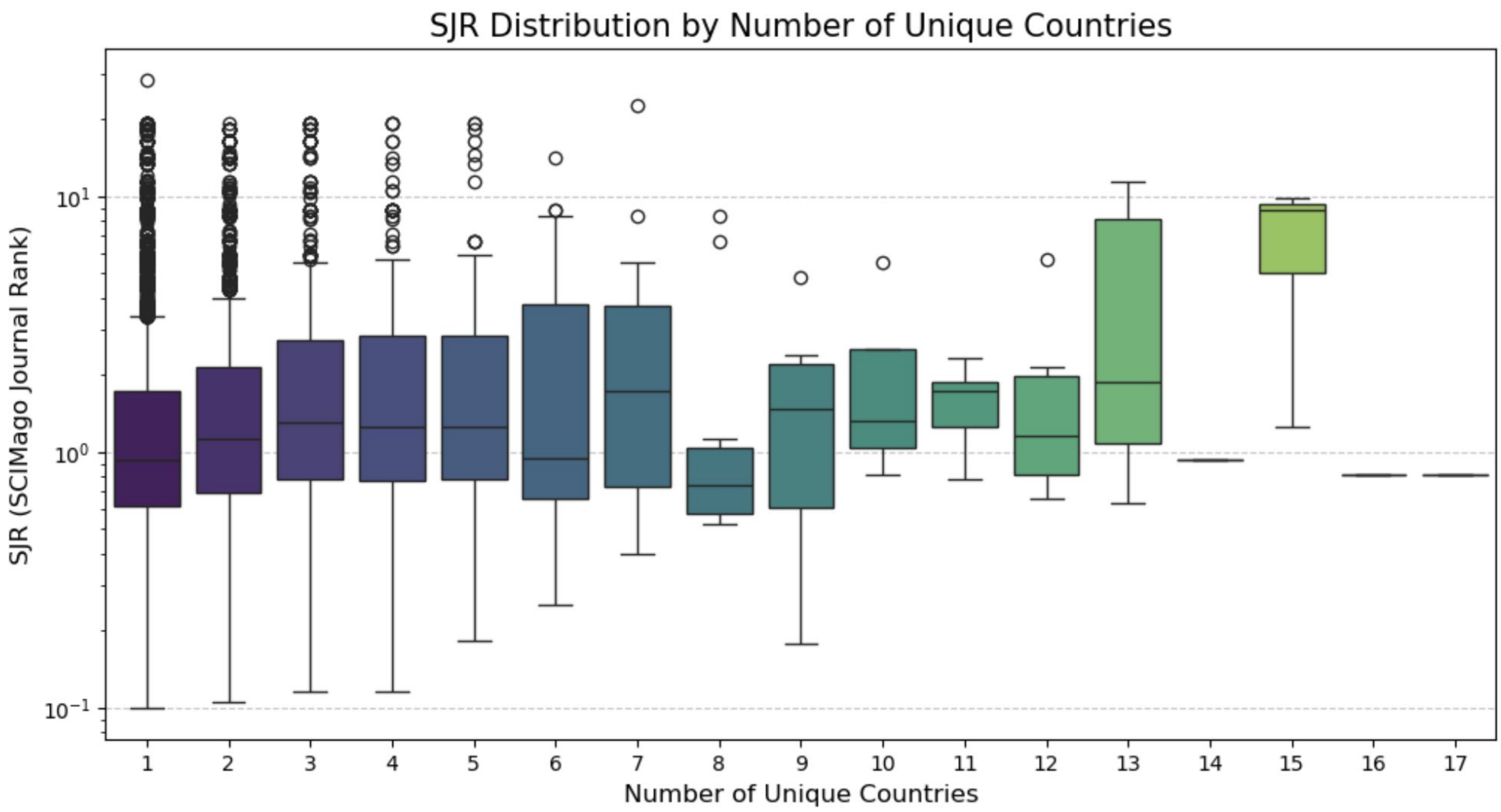
- Reflects the scientific prestige and influence of a journal based on citations.
- Higher SJR values represent greater prestige, considering both the number of citations and the prestige of the citing journals.
- Often used to rank journals within specific subject areas.


FWCI Distribution by Number of Unique Countries
FWCI (Field-Weighted Citation Impact)
10^2
10^1
10^0
10^{-1}
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
Number of Unique Countries

International Collaboration Premium: How Country Diversity Impacts FWCI

1. Correlation Between Diversity and Impact

- Generally, as number of unique countries increases, Mean and Median FWCI tend to rise.
- Why? Papers benefit from wider networks, diverse expertise, and higher visibility across geographic regions.
- The 'Global Reach' Effect: A paper with 5 countries (e.g., 🇺🇸, 🇬🇧, 🇨🇳, 🇧🇷, 🇩🇪) is more likely cited in all five regions compared to a single-country paper.

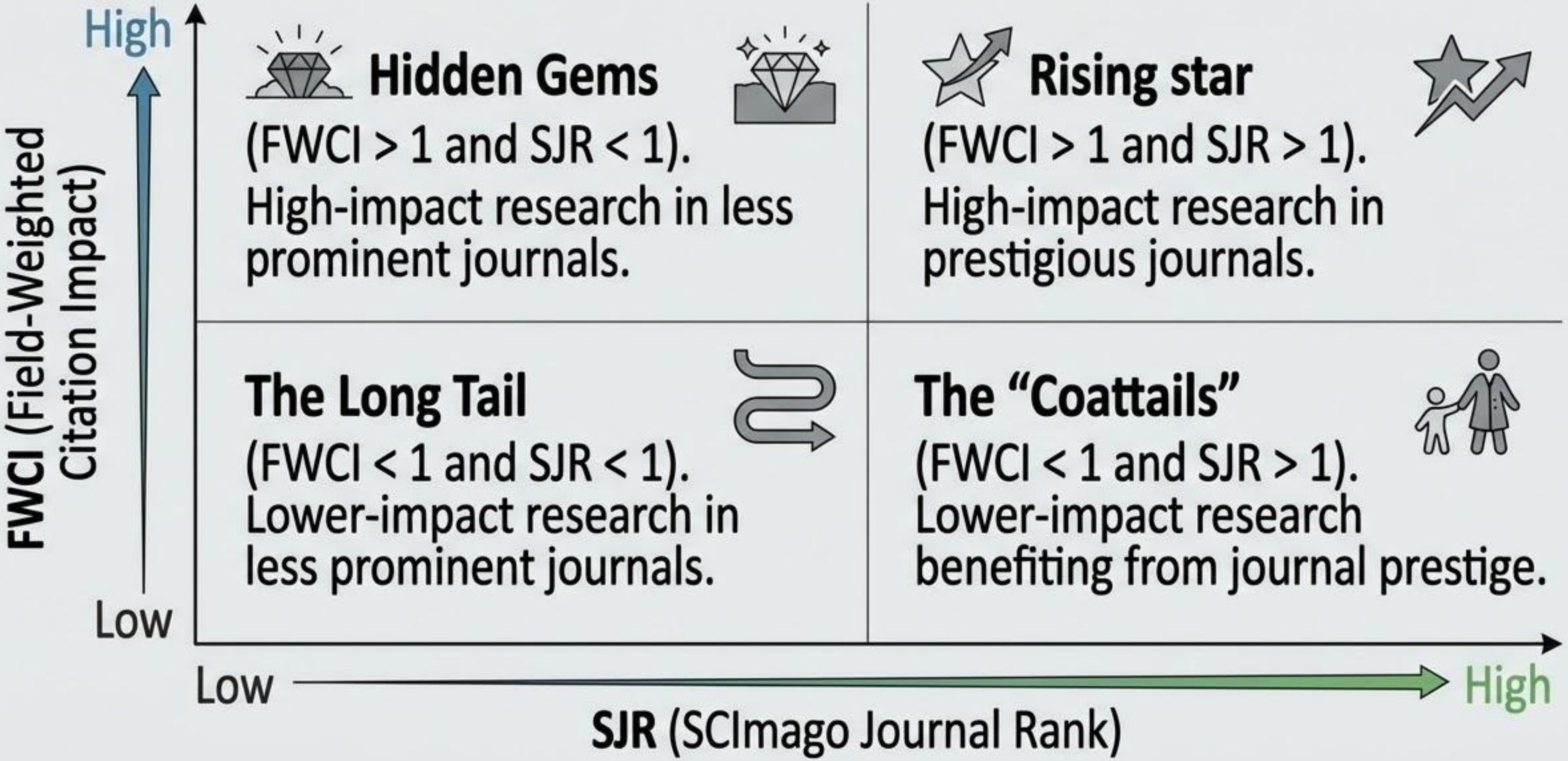

SJR Distribution by Number of Unique Countries
SJR (SCIMago Journal Rank)
10^1
10^0
10^{-1}
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
Number of Unique Countries

SJR Distribution by Number of Unique Countries

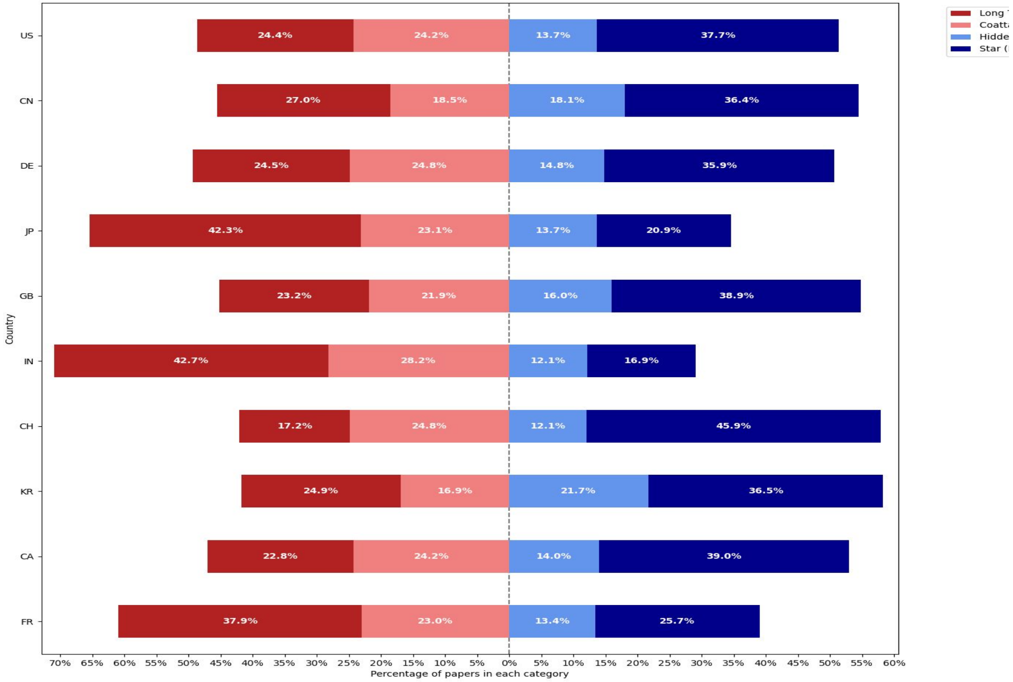
Key Insights

- **Clear Trend:** Higher country diversity correlates with higher median SJR.
- **Multi-national teams** are more successful in publishing in top-quartile (Q1) journals.
- **'Prestige Floor':** High collaboration groups have fewer low-impact venues.
- **Single-country** papers show a wider, more volatile range of journal quality.

Journal Classification Matrix: FWCI vs. SJR



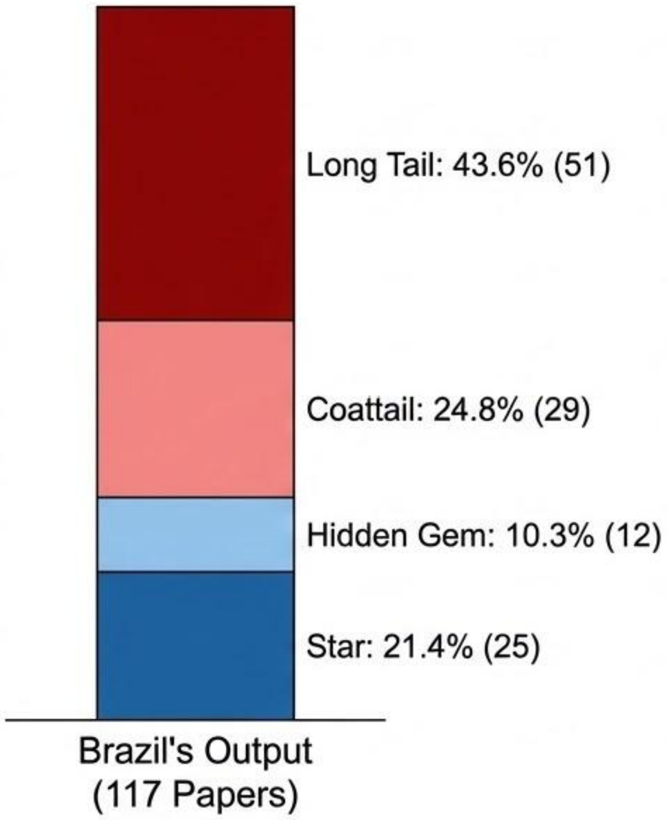
Quality indicators by country



Highlights: Quality Indicators by Country

- **China (CH)** leads in high-quality "Star" publications (**45.9%**) and has the lowest proportion of "Long Tail" papers (**17.2%**).
- The **United States (US)** and **United Kingdom (GB)** also demonstrate strong performance with high percentages of "Star" publications (**37.7%** and **38.9%**, respectively).
- **India (IN)** has the highest proportion of "Long Tail" publications (**42.7%**) and the lowest percentage of "Star" papers (**16.9%**), indicating a challenge in publication quality.
- **Japan (JP)** shows a similar profile to India, with a high "Long Tail" (**42.3%**) and relatively low "Star" percentage (**20.9%**).
- **South Korea (KR)** is notable for having the highest proportion of "Hidden Gems" (**21.7%**), suggesting high-impact research in less prominent journals.

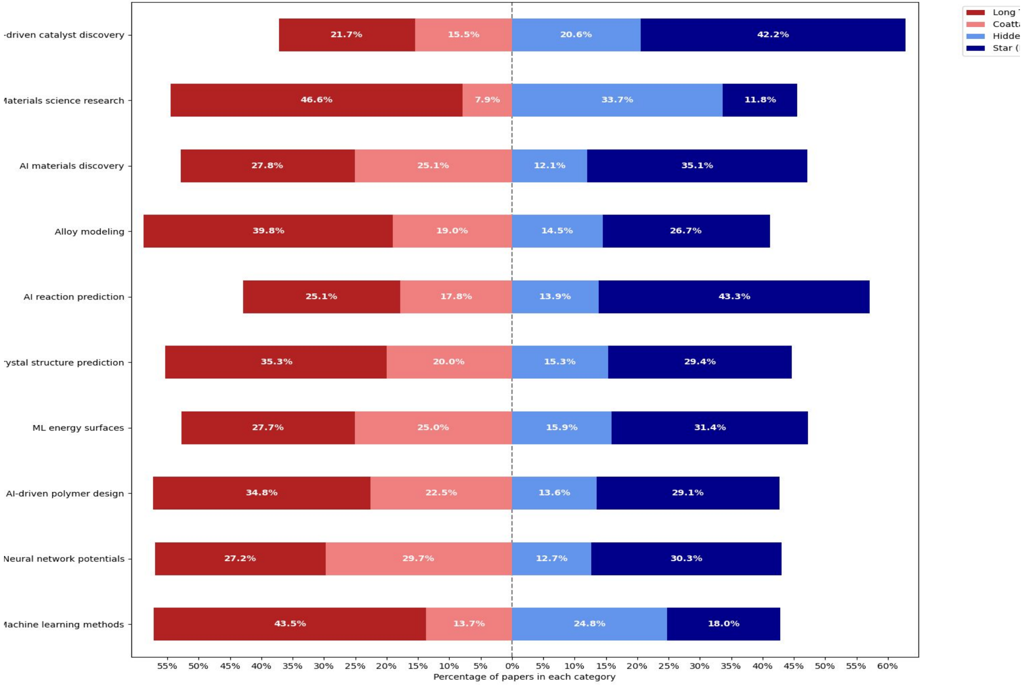
Brazil's Research Profile



Key Comparisons & Summary

- **High Concentration in 'Long Tail' (~44%)**: Significant portion in lower-tier journals. Nearly double the US (~24%) or China (~27%). Similar to India and Russia.
- **Gap in 'Star' Papers (~21%)**: Lower than leaders like the US (~38%) and China (~36%), and volume peers like Belgium (41%).
- **'Coattail' Rates (~25%) are Globally Consistent**: Similar to the US, Russia, and Denmark, indicating a universal challenge in converting prestigious placements into high citations.
- **Summary Assessment**: Represents a volume-driven, developing research ecosystem, similar to India and Russia, but distinct from highly efficient European nations.

Quality index by topic



Highlights: Quality Index by Topic

- ⭐ **AI-Driven Topics Lead in "Star" Papers:** Topics like "AI reaction prediction" and "AI-driven catalyst discovery" show the highest proportion of "Star" papers (over 40%), indicating high impact.
- 📉 **Broader Topics Have Longer "Tails":** General categories like "Materials science research" and "Machine learning methods" have a larger share of "Long Tail" papers, suggesting a wider range of quality.
- ⚖️ **Modeling and Prediction Show Balance:** Topics such as "Alloy modeling" and "Crystal structure prediction" exhibit a more even distribution across all quality categories.
- 🔍 **Niche AI Applications Yield "Hidden Gems":** Specialized areas like "AI-driven polymer design" and "Neural network potentials" have a significant portion of "Hidden Gem" papers, highlighting their potential.

LLMs for Materials: The Quality Index Leader

Highest-performing topic in impact & prestige, drastically outperforming global averages.

1. The Undisputed Leader in "Star" Papers & Shrinking "Long Tail"

⭐ **47.9%** (LLMs) vs. 30.8% (Global Avg.) Star Papers. Rank #1/44 topics for highest concentration.

📉 **18.6%** (LLMs) vs. 31.0% Rare (Global Avg.) Long Tail Papers. publication in low-impact venues.

3. Comparison Table: LLMs vs. Global & Other Topics

Topic	⭐ Star (Gold)	💎 Hidden Gem	Coattail (Pink)	📉 Long Tail
LLMs for materials	**47.9%**	**14.5%**	**19.0%**	**18.6%**
AI-driven catalyst discovery	**42.2%**	20.6%	15.5%	**21.7%**
GLOBAL AVERAGE	**30.8%**	**17.5%**	**20.6%**	**31.0%**
Materials science research	**11.2%**	33.7%	16.5%	**38.6%**

Summary Conclusion

The data reveals a compounding 'gold rush' effect. Fast volume growth combined with aggressive community reward (high acceptance & citations).

Conclusion: The Power of Bibliometric Insight

- 🔍 **Bibliometric analysis** is crucial for mapping scientific landscapes, identifying emerging trends, and understanding research impact.
- Data reveals a significant shift towards AI-driven research, with distinct quality profiles across different topics, as seen in the 'Star' and 'Long Tail' distributions.
- 🚀 **Specialized fields** like 'LLMs for Materials" are experiencing explosive growth and high community recognition, victinining
growth and high community recognition, highlighting a 'gold rush' effect in science.

Leveraging these insights enables strategic decision-making in research and development.

Thank You!

- Thank you to PhD student Luis Cesar Azevendo
- Grateful for the support from the CINE CMD team for the valuable contributions and discussion provided
- Presentation made possible through this collaboration