📌 Scinapse FAQ Contents

📌 Data coverage & Processing

Q1. From where does Scinapse aggregate data (what kinds of papers and articles does Scinapse handle)?

Scinapse utilizes Pluto Labs' proprietary web spider crawling technology to collect and index content from the vast majority of publicly accessible journal websites across the web. (ScienceDirect, SpringerOpen, Wiley Online Library, IEEE Xplore etc.) Scinapse aggregates scholarly literature from a broad range of **open and publicly accessible sources**, covering journal articles, conference papers, and preprints across STEM fields.

Its primary data sources include major open scholarly databases such as:

- Microsoft Academic Graph
- PubMed and PubMed Central
- OpenAlex
- Semantic Scholar

As a result, Scinapse handles a wide spectrum of academic outputs, including journals, conferences, and preprints.

Q2. How does Scinapse collect, select, and update its data? How is it different from Web of Science (WoS) and Scopus?

Web of Science and Scopus curate a fixed list of journals based on their own selection criteria and receive bibliographic data directly from publishers. In contrast, Scinapse does **not** maintain a pre-selected journal list. Instead, it collects as many papers as possible from both the open web and publicly accessible databases.

In terms of data processing, both Scopus and Scinapse employ AI-based methods for author and affiliation disambiguation, whereas WoS relies more heavily on traditional rule-based and text-matching approaches.

The most significant difference lies in how citation data is constructed and used:

- **WoS and Scopus** provide a fixed citation graph limited to the papers included in their databases.
- **Scinapse** allows users to dynamically define and adjust the citation set depending on their research or analysis purpose.

This flexibility enables Scinapse to:

- Deliver **faster and more contextually accurate discovery** of relevant papers
- Support **adaptable, purpose-driven citation analysis** tailored to specific research domain

One trade-off is that certain traditional library-oriented metadata (such as page numbers) may be less complete; however, these are no longer essential in modern web-based research workflows.

Q3. Does Scinapse cover conference papers?

Yes, Scinapse covers conference papers.

We index the **entire arXiv repository,** which is widely used in conference-driven research fields—most notably Computer Science, where many papers are shared on arXiv prior to conference publication. Therefore, in fields like **Computer Science**, **Scinapse provides much better coverage**, unlike WoS or Scopus, which tend to miss a large number of such papers.

While conference classification is not yet perfect, it is more than sufficient for identifying and analyzing research trends. We also plan to further enhance conference coverage by integrating additional data sources such as DBLP and Semantic Scholar.

Q4. If Scinapse relies mainly on open or publicly available papers, how does this affect cases where relevant papers are behind a paywall or not available online?

If a paper is not available online, Scinapse cannot directly access its full text.

However, this does **not** pose a meaningful limitation in modern research workflows:

- Research discovery primarily depends on **bibliographic metadata, citation relationships, and abstracts**
- These signals are sufficient for identifying reliable and influential work, regardless of paywall status
- Even when full text is available, Scinapse often relies on structured metadata and citation information for efficiency and robustness

From a research perspective, the goal is not merely to find *any* relevant paper, but to identify **reliable, influential research**, which Scinapse achieves effectively without requiring full-text access in all cases.

Q5. What is the time range of Scinapse's database? Can users filter by publication year?

Scinapse can ingest papers from **any publication year**, provided that:

- The paper has a DOI
- The paper has been digitized and made publicly accessible by the publisher

The coverage range is therefore determined by **publisher digitization**, not by limitations in Scinapse's data collection.

Scinapse also provides a **search-by-time filter**, allowing users to restrict analyses to specific publication periods (e.g., last 5 years, historical trends, etc.).

Q6. What is the “Citation Impact” shown in the Journal tab? Is it the same as JIF?

Yes. The “Citation Impact” displayed in the Journal tab represents the same value as the Journal Impact Factor (JIF).

Scinapse collects publicly available JIF values from each journal’s official website.
Since journals update their websites *after* WoS releases JIF, there may be a **3–5 week delay** before Scinapse reflects the latest numbers.

Aside from this delay, Scinapse’s JIF values correspond to **WoS 2-year JIF**.

Q7. Where does Scinapse obtain affiliation information? How does it compare with ORCID?

Affiliation information comes directly from publication metadata, which means each paper reflects the author’s affiliation **at the time the paper was written**.

For convenience, Scinapse also provides a **latest affiliation** on each author’s profile.
Individual paper pages show the affiliation relevant to that specific publication.

In contrast, ORCID depends on authors manually updating their information, so the latest data may not always be reflected.

Q8. Is it possible to distinguish authors with the same name in Scinapse?

Yes, Scinapse can accurately distinguish between authors with the same name.

Traditional academic databases typically rely on basic information such as names to identify authors, which makes it difficult to accurately differentiate individuals with identical names. In addition, many of these systems were built decades ago, making it challenging to apply modern AI technologies and often requiring manual corrections.

In contrast, Scinapse uses AI-based author disambiguation (clustering) technology to identify authors based on a rich set of research signals, including:

- research topics
- citation patterns
- co-author networks

For example, even if there are many researchers with the same name, Scinapse can distinguish them by analyzing their research areas and collaboration networks, recognizing each as a unique individual.

Because this process is fully automated and continuously updated, author profiles remain accurate and up to date.

This allows researchers to quickly find precise publication records and research outputs for each author in Scinapse.

📌 Data Quality & Result Validity

Q9. How are research topics for papers and authors classified? How does the system determine whether a paper or author belongs to a specific field or keyword?

Scinapse analyzes research topics by combining **deep learning–based field classification** with **keyword matching**.

First, **field classification** is performed using a **deep learning** model. It comprehensively analyzes a paper's title, abstract, and citation relationships to automatically classify its research area based on context, rather than relying on simple keyword matching.

Then, within the identified field, **keyword matching** is applied to precisely connect relevant papers and authors.

Compared to simple keyword-based search, this approach enables:

- **More accurate topic classification by incorporating context**
- **Faster and more relevant results aligned with the user's research interests**

Q10. Scinapse collects large amounts of data. How are inaccurate or low-quality entries handled?

Scinapse does not manually inspect or censor individual papers for accuracy.
However, papers lacking essential metadata (such as DOI) are excluded to ensure indexing stability.

We rely on **citation patterns and natural academic ecosystem signals** to reflect paper reliability—i.e., the research community determines influence through citations.

Scinapse's open, researcher-centric approach differs from the closed journal-list methodology of traditional databases.

Q11. How does Scinapse ensure clean, credible data and prevent hallucinations in AI outputs (e.g., reviews, analyses)?

Scinapse utilizes AI designed to fundamentally avoid the main cause of hallucinations commonly observed in general-purpose LLMs.

Most LLM-based tools generate responses by predicting tokens solely based on their training data, which can lead to fabricated or unverified information.

Scinapse takes a fundamentally different approach:

- **The LLM is not allowed to search for or evaluate papers on its own**
- All paper retrieval, filtering, and evaluation are performed by Scinapse's **internal search and citation-analysis engine**
- The LLM's role is strictly limited to **summarizing and structuring information that has already been validated**

Because the LLM only operates on a **pre-curated, citation-validated paper set**, Scinapse ensures that outputs are grounded in clean, credible, and traceable academic data.

This architecture significantly reduces hallucinations and ensures research-grade reliability.

📌 Metrics Calculation & Definitions

Q12. How is h-index calculated? Is it the same as in WoS or Scopus?

Scinapse recalculates the h-index based on its own internal database.
Scinpase presents two kinds of h-index: Total h-index and domain h-index.

The total h-index is computed automatically from the author's full publication list. Conceptually it is the same as WoS/Scopus, but the actual numbers differ because each platform has a different underlying dataset.(Even WoS and Scopus do not match each other.)

Domain-specific h-index values are **dynamically recalculated** depending on the domain selected. (Further details are provided in the next section.)

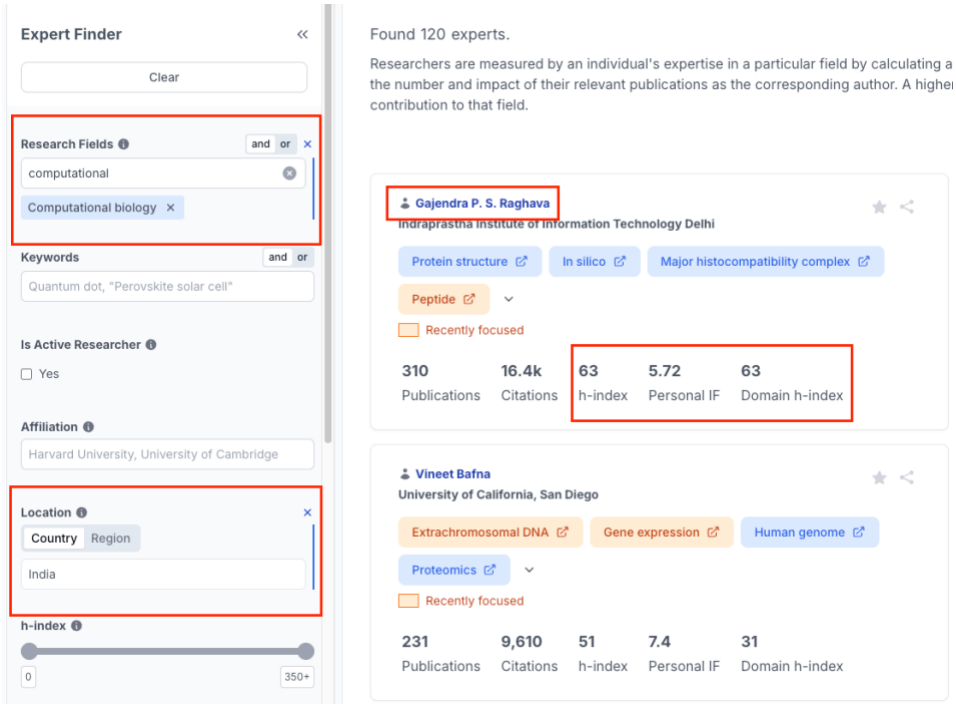
Q13. What is Domain h-index?

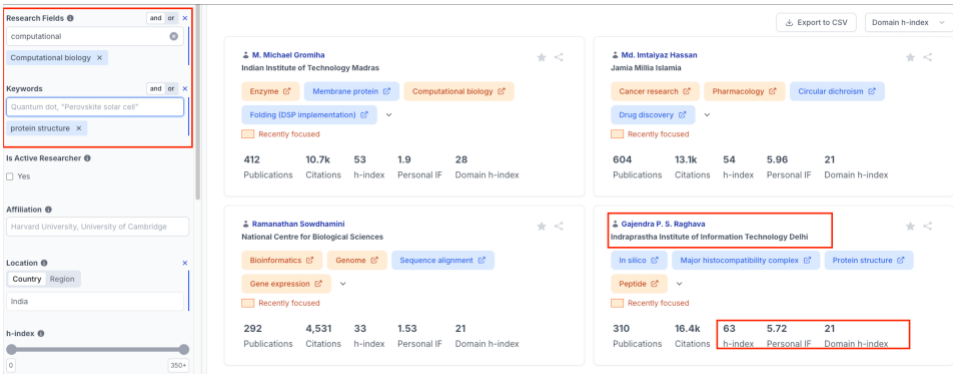
Domain h-index is calculated from a filtered set of papers specified by the user, considering **only the papers for which the researcher is the first author or corresponding author** within that domain.

In other words, instead of using all publications, this metric focuses solely on:

- papers belonging to a specific research field (filtered set), and
- papers where the researcher played a *leading* authorship role (first or corresponding author).

Domain h-index provides a more accurate representation of the researcher's true influence within the specific field the user is examining, because it takes into account whether the researcher was a primary author.

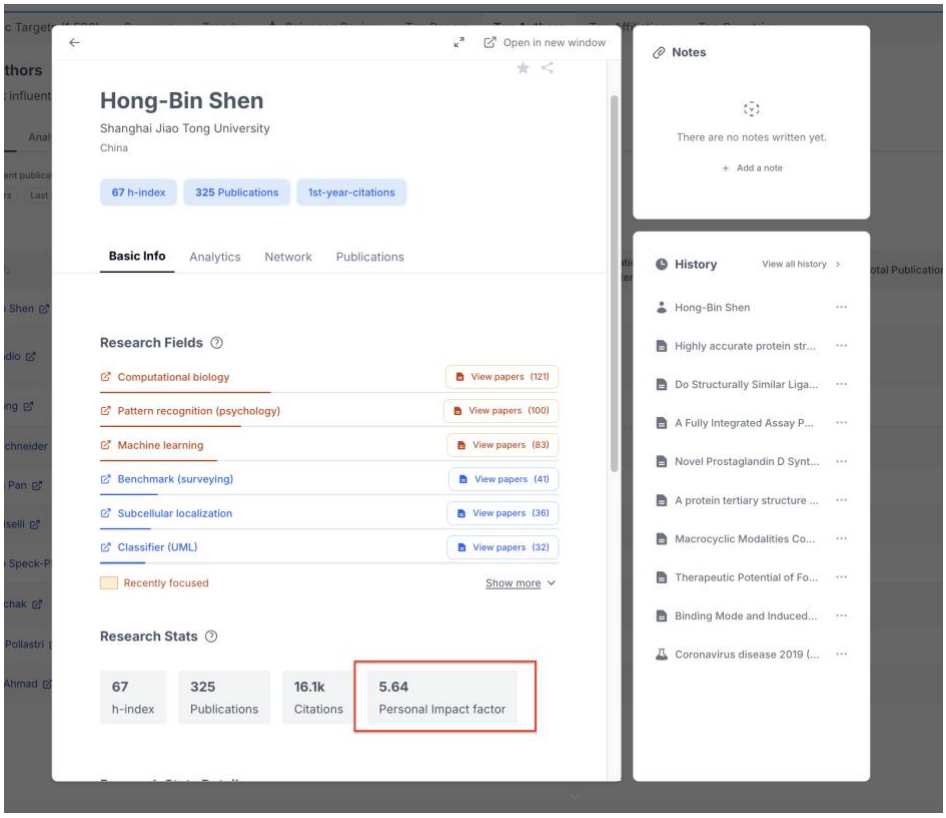




For example, if the **Domain h-index is approximately equal to the overall h-index**, this indicates that: “most of the researcher’s citation impact comes from a particular domain, and the key papers in that domain were led by the researcher.”

**From the two screenshots above, you can see that the author’s Domain h-index changes when additional keywords are applied and the filtered set is updated.*

Q14. What is Personal IF?



Personal Impact Factor (Personal IF) indicated in the author tap is an informal metric that calculates the average impact of a researcher's recent papers using the same method as a journal's Impact Factor. It's an internal metric to assess author's impact, compensating h-index.

To do this, we **group all papers the researcher published in the past two years into a single set**, as if they were issues of one "personal journal" and then apply the Journal Impact Factor formula to that set.

Although not an official bibliometric indicator, it provides a helpful way to assess the average quality of a researcher's publications.

The interpretation is straightforward:
if a researcher has a Personal IF of 34.27, it suggests that "the average influence of the papers he/she published in the past two years is comparable to papers typically found in a journal with an Impact Factor of around 34."

In summary: Personal IF reflects the average impact of papers authored by an individual researcher.

📌 Main Features: Purpose and Use Cases

Q15. Why does Scinapse include Research Intelligence, Find Expert, and Trends features, and how are they used?

Scinapse includes **Research Intelligence, Find Expert, and Trends** as part of its core functionality to help researchers **dramatically reduce literature review time** by improving both **search accuracy** and **analytical depth**.

Many researchers, particularly in science and engineering, do not actively rely on traditional discovery tools such as Web of Science or Scopus for everyday research exploration. This is because conventional catalog-style systems often struggle in real research workflows, where:

- A single keyword search can return thousands of loosely related papers
- Author databases do not distinguish actual research contribution, making it difficult to identify true experts

Scinapse was designed to address these limitations by offering the following core analytical features:

- **Research Intelligence (RI):** Quickly identifies research dynamics around a specific topic
- **Find Expert (FE):** Identifies influential researchers in a field
- **Trends:** Helps assess research activity before choosing a research direction

---

Q16. Can Scinapse generate reviews filtered by a specific institution or affiliation (e.g., to assess research progress within an institute)?

Yes. Scinapse fully supports affiliation-based filtering.

In the Research Intelligence tab Users can:

- Generate reviews limited to papers authored by researchers from a **specific institution**
- Further restrict results to **first-author or corresponding-author papers**
- Combine affiliation filters with **domain, keyword, and time-range filters**

This makes it possible to analyze institutional research performance, progress in a specific field, or strategic research focus areas with high precision.

Q17. Can Scinapse identify key opinion leaders (KOLs) and compare institutional research performance using both quantitative and qualitative metrics?

Yes.

Scinapse supports multiple approaches to identifying KOLs and evaluating research performance:

- Author contribution levels (first/corresponding author)
- Citation-based metrics
- Domain-specific influence analysis
- Qualitative paper-level insights generated by Scinapse's AI model

A key differentiator is Scinapse's ability to generate **field-specific author expertise scores**, allowing users to identify true domain leaders rather than broadly cited authors.

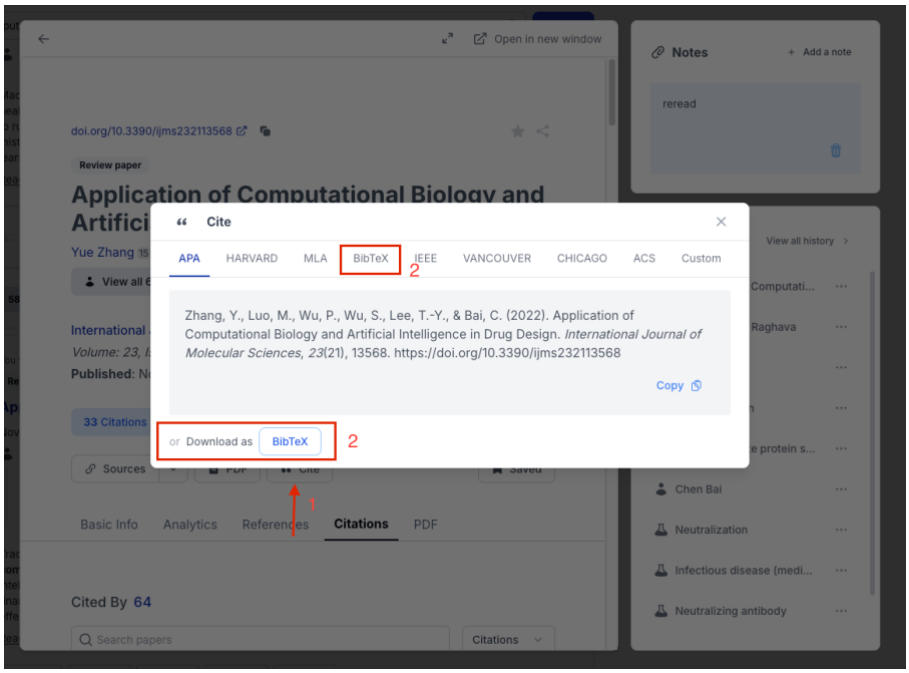
This enables meaningful comparisons between individual researchers, institutions, or research groups using both **quantitative indicators and contextual scientific relevance**.

Management & Institutional Functions

Q18. Citation export and EndNote compatibility

EndNote integration is supported.

Citations can be exported in **BibTeX** format, which can be directly imported into EndNote.



You can find the BibTeX export option by opening the *Article* tab, clicking **"Cite"**, and then downloading the BibTeX file. You can access the **BibTeX tab** and copy the generated URL or code.

Q19. Can users access Scinapse from off-campus (external IP)?

Yes. Scinapse supports several methods for off-campus access:

1) IP Range Authentication (On-Campus Automatic Access)

For on-campus users, Scinapse also supports **IP-based authentication**.

2) 90-Day External Authentication

Scinapse provides seamless off-campus access through a 90-day external authentication system. When a user signs up (creates an account) **while connected to the university's on-campus IP range** (including campus Wi-Fi):

- Their account becomes authorized for off-campus access for **90 days**
- During this period, they can access Scinapse from any external network
- To extend access, the user simply reconnects to the campus network once every 90 days

This allows users to work continuously without relying on a proxy system.

3) Proxy Access

If the university operates a proxy system, Scinapse can register the library's proxy IP address

---

Q20. Can institutional administrators download usage reports directly?

Currently, usage reports are generated by our team upon request.
However, we plan to introduce an **Admin Dashboard** that allows institutional administrators to download usage reports directly.

Q21. Does Scinapse provide an API?

Yes, Scinapse provides an API.