

AI Research Lens

Making AI-assisted research visible, verifiable and teachable



FROM RESEARCH QUESTIONS TO INSPECTABLE EVIDENCE

A researcher-centred, AI-augmented workspace for bibliometric analysis, custom computational exploration, open-access paper discovery and evidence-grounded interpretation.

Adela Bâra, Simona-Vasilica Oprea

Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, Romania

bara.adela@ie.ase.ro; simona.oprea@csie.ase.ro

Why AI Research Lens?

Modern research is rich in scholarly data but fragmented across bibliometric tools, scripting environments and AI assistants. Researchers often need specialist programming skills to move from descriptive mapping to custom indicators, and generated answers can be difficult to trace back to the underlying evidence.

The objective

AI Research Lens connects established scientometric methods with inspectable generative-AI functions in a single research workflow. It helps researchers explore bibliographic collections, formulate new analytical questions and ground literature synthesis in selected full texts, without treating the model as a substitute for scientific judgement.

Five design priorities

- **Accessibility** - Express an analytical goal in natural language.
- **Analytical flexibility** - Go beyond fixed dashboards to tailored metrics, plots and exploratory comparisons.
- **Transparency** - Keep data, analysis scripts, intermediate results and document references visible.
- **Connected exploration** - Use detected topics and keywords to drive targeted open-access paper discovery.
- **Research education** - Make methods, evidence selection, verification and limitations explicit learning activities.

THE DIFFERENCE

AI Research Lens is not simply “ask a chatbot about a dataset”. The application separates dataset-derived computation from language-model interpretation, allows user to query the bibliometric dataset in natural language and transpose the query into an executable Python code, exposes the generated code for review, and provides a route from corpus-level patterns to page-referenced document evidence.

How the system works

AI Research Lens uses a modular architecture that combines a data layer, a deterministic analytics layer, an AI-agent layer and an interactive user interface. Four complementary AI functions extend the bibliometric engine.

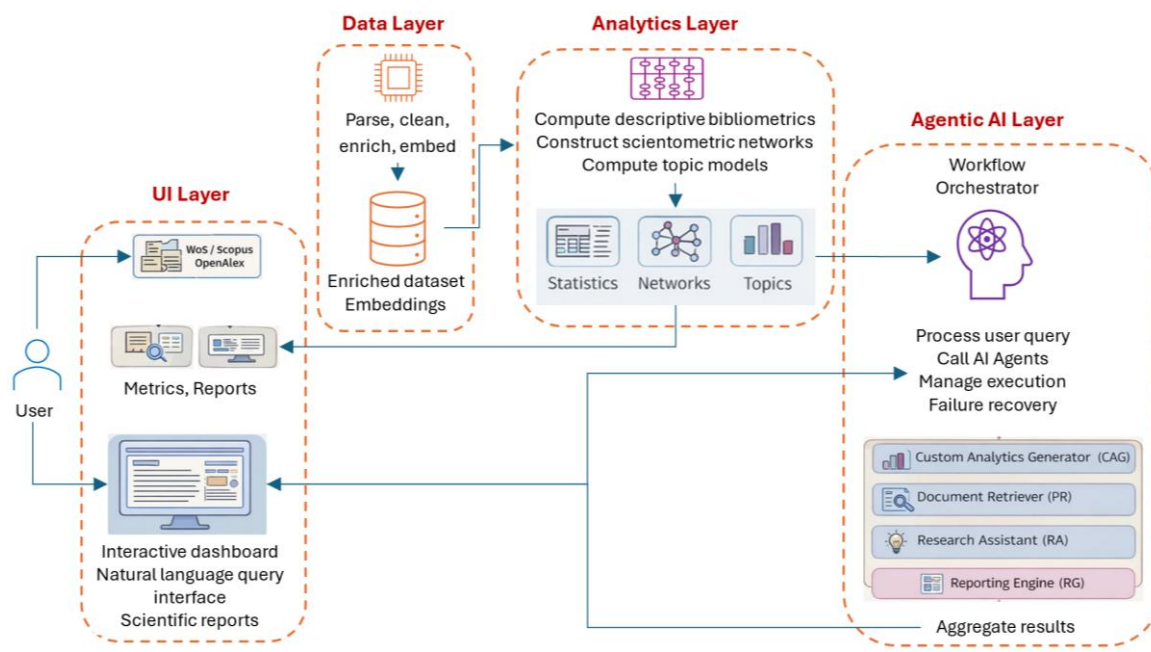


Figure 1. Four-layer architecture of AI Research Lens

Inside the methodology: the four layers

Each layer has a distinct responsibility, allowing classical scientometric computation to remain distinguishable from AI-generated code and interpretation.

1. Data layer — turn records into a research-ready corpus

Bibliographic records extracted from WoS, Scopus or OpenAlex are imported and harmonised into common fields for publications, authors, affiliations, countries, years, abstracts, keywords and references. The proposed processing pipeline includes author and institution normalisation, DOI-based and fuzzy duplicate matching, keyword consolidation, text cleaning and reference parsing. Text enrichment extends this metadata foundation through document embeddings, semantic keyword expansion and entity extraction. The enriched corpus can feed clustering, network construction, semantic search and later agent reasoning.

2. Analytics layer — compute the structure of a field

- **Descriptive bibliometrics** - Publication growth, journals and countries, citation indicators, author profiles and keyword frequency distributions.
- **Network analytics** - Co-authorship, institutional collaborations, keyword co-occurrence, co-citation and bibliographic coupling; where appropriate, measures of centrality and network communities.
- **Topic modelling** - LDA and BERTopic for identifying themes, topic hierarchies, coherence and evolution over time.

These modules generate dataset-derived indicators, networks and figures that can be inspected before any language-model synthesis.

3. Agentic AI layer — four specialist agents

- **Custom Analytics Generator (CAG)** - Interprets a natural-language analytical query, produces Python tailored to the loaded dataset, checks the resulting code and lets the researcher inspect and edit it before execution.
- **Paper Retriever (PR)** - Expands research themes into searches, discovers open-access works through OpenAlex, downloads available PDFs and extracts text for the evidence corpus.

- **Research Assistant (RA)** - Uses retrieval-augmented generation to interpret relevant full-text passages, explore trends, compare methodologies and investigate gaps, with page-level references for source checking.
- **Report Generator (RG)** - Combines available metrics, figures, tables and interpretations into a coherent report and export them to PDF/HTML research summaries.

The workflow orchestrator — coordination, not a fifth AI agent

A deterministic orchestrator connects the four agents and analytics modules. Its defined tasks are dependency management, shared session state, passing intermediate artefacts between components, validation, failure handling and routing results to the interface or reporting stage. It supports iterative research rather than isolated chatbot responses.

4. User interface layer — two ways of working

The graphical interface supports corpus upload and inspection, standard descriptive/network/topic analyses, PDF selection and report viewing. Natural-language interaction complements these controls when a researcher needs a non-standard analysis or source-grounded explanation. Both interaction paths are designed to retain visible inputs and outputs.

Application insights

1. Prepare and understand the corpus

The application loads spreadsheet or text records, currently centred on Web of Science-style metadata, and prepares publication years, authors, keywords, citations and other bibliographic fields for analysis.

What a researcher can inspect

- **Data quality** - Available columns, record count, publication span and bibliographic field coverage.
- **Core statistics** - Publication volume, citations, unique authors and keyword vocabulary.
- **Temporal development** - Publication trends and citation patterns across years.

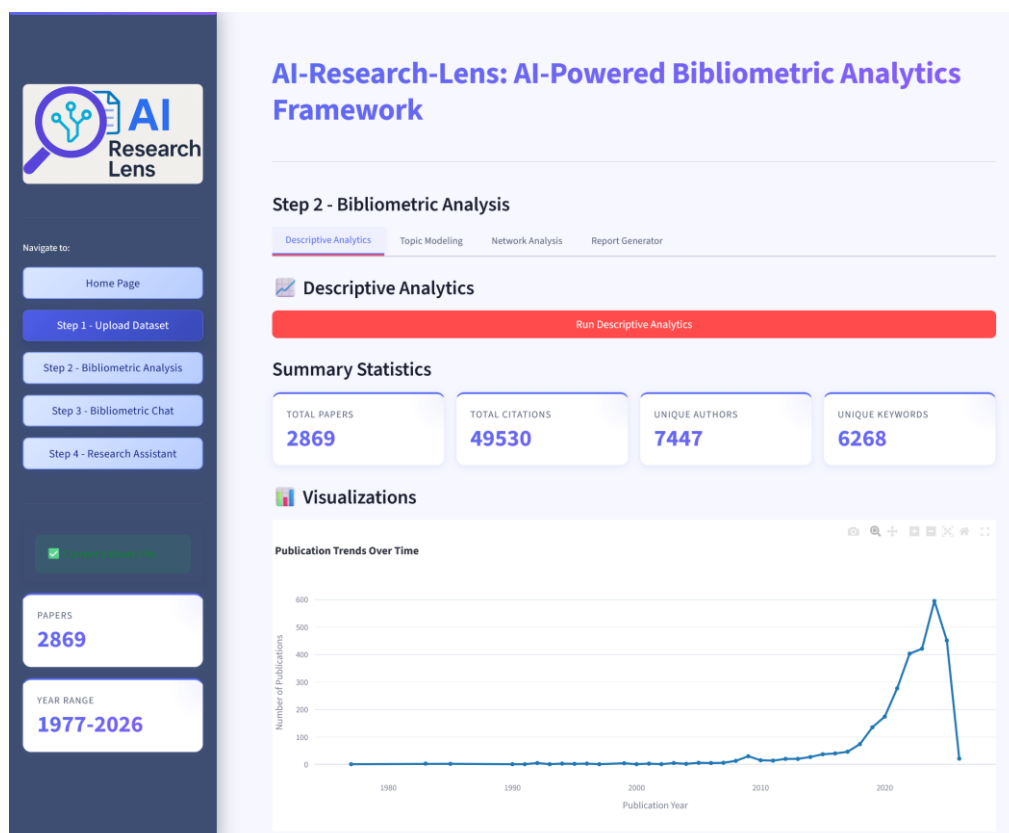


Figure 2. Application screenshot. Descriptive statistics generated for the bundled energy-communities demonstration corpus.

2. Map the research landscape

The analytics workspace combines familiar bibliometric indicators with advanced thematic and relational views. Researchers can move from aggregate trends to topics, collaborative structures and emerging areas of enquiry.

- **Descriptive bibliometrics** - Publication and citation trends, leading sources, authors, affiliations and keyword distributions.
- **Topic modelling** - BERTopic and LDA for research-theme discovery, topic relationships and thematic interpretation.
- **Network analysis** - Co-authorship, keyword co-occurrence, citation-oriented relationships and network centrality measures.

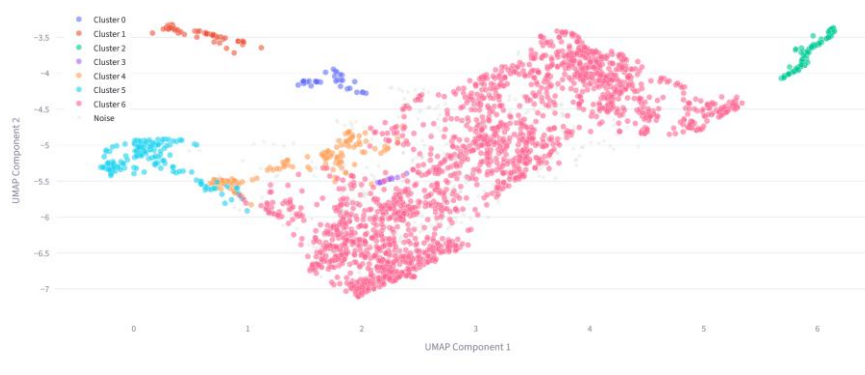


Figure 3. Example embedding-based thematic visualisation from the research study.

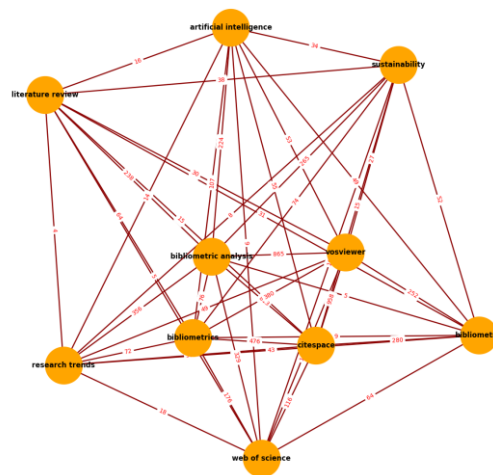


Figure 4. Example research collaboration network from the research study.

These figures illustrate data-analysis outputs. Their scientific meaning depends on the corpus, preprocessing decisions and model parameters selected.

3. Create your own analysis with AI

A researcher types a question in ordinary language and the Custom Analytics Generator generated Python code for a dataset-specific operation. The proposed computation is visible and editable, and execution is a separate action controlled by the user.

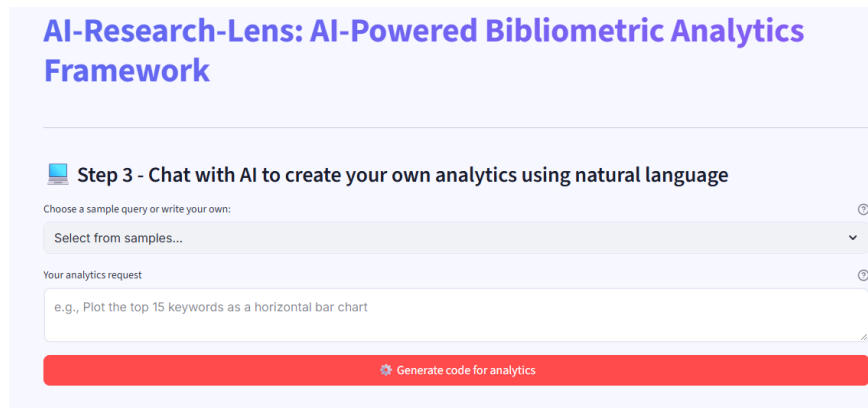


Figure 5. Application screenshot. The natural-language interface for generating custom analytics.

Example research instruction

User query: “Compare annual publication growth before and after 2018, visualise the trend and identify the years with the largest changes.”

Processing steps:

- **Generate** - Translate the research objective and dataset context into a candidate script.
- **Review** - Inspect the generated imports, filters, grouping logic and intended outputs.
- **Execute and interpret** - Run the analysis and compare numerical results, plots and optional AI commentary.

This is useful in research-methods education since students can assess not only the answer, but whether the code actually operationalises the research question.

4. From field mapping to full-text evidence

The Paper Retriever uses terms discovered during bibliometric analysis to suggest OpenAlex searches. Available open-access PDFs can be processed into a searchable evidence base, allowing the researcher to move beyond titles, abstracts and citation counts.

Step 4 - Research Assistant

Paper Retriever Ask Research Assistant

Retrieve open access papers from OpenAlex

▼ Suggested searches derived from your bibliometric analysis

These queries are derived from your BERTopic clusters, top keywords, and network centrality scores. Click any to use it as your search query:

demand response	electric vehicle	energy management
energy sharing	energy storage	energy transition
local energy community	optimization	prosumer

Search for papers containing the following keywords: ?

e.g., fake news detection machine learning

Max Papers ?

10 - +

> Advanced Filters (Optional)

Search for Papers

Figure 6. Application screenshot. Topic- and keyword-guided paper discovery.

5. Research Assistant: evidence before synthesis

For question answering, the application chunks document text, combines semantic and lexical retrieval, can rerank passages, and supplies selected context to the language model. Answers are instructed to remain within retrieved evidence and end with document page references.

User asks an evidence-based question: “What methods and datasets were used in these papers? Which pages support each part of the answer?”

Page references allow the researcher to verify claims against the original documents. Missing evidence, incomplete PDF extraction and imperfect retrieval remain important limitations.

An AI-ready assistant for research in universities

AI Research Lens is designed as a research assistant, and its value is the possibility of making decisions along a research workflow visible and discussable rather than presenting AI output as a final, opaque product.

Where it could be used

- **Research-methods courses** - Teach corpus selection, bibliometric interpretation, code inspection and responsible use of AI.
- **Doctoral training** - Support structured literature exploration, topic discovery and reproducible custom analyses.
- **Library and research support** - Demonstrate transparent search, analytics-guided discovery and source verification.
- **Interdisciplinary research** - Explore collaboration networks and emerging themes across large publication collections.