Research Knowledge Navigator: AI-Powered Research Intelligence Platform

Core Idea:

This is a web application that ingests research papers (PDFs, arXiv, PubMed, IEEE), builds a dynamic knowledge graph of methods, datasets, authors, and institutions, and provides multi-audience explanations, trend forecasts, and interactive exploration tools. Unlike generic chatbots, it is a full research intelligence system that organizes, explains, predicts, and teaches.

Features

1. **Smart Corpus Management**

   o Upload multiple research papers at once.

   o Auto-cluster papers by topic for easy organization.

   o Enable semantic and entity-based search for methods, datasets, and concepts.

2. **Concept Dependency Maps**

   o Generate prerequisite maps as directed acyclic graphs (DAGs).

   o Example: RNN → LSTM → Attention → Transformer.

   o Helps students and researchers understand learning pathways and study sequences.

3. **Evidence-Backed Q&A**

   o Retrieval-Augmented Generation (RAG) over knowledge graph + paper content.

   o Every answer cites exact source passages for transparency and credibility.

4. **Analytics Dashboard**

   o Track trending methods, datasets, and institutions.

   o Display heatmaps showing dataset-method co-occurrence.

- Show collaboration networks and leaderboards (e.g., top-cited methods in a domain).

5. **Trend Forecasting**

   - Time-series analysis of method/dataset popularity.

   - Predict emerging techniques and visualize research timelines.

   - Example: CNNs peaked in 2019 → Transformers overtook in 2021 → Diffusion models trending now.

6. **Collaboration Recommender**

   - Suggest potential collaborators using graph centrality and co-authorship analysis.

   - Example: "Authors X and Y work in complementary domains but haven't collaborated yet."

7. **Learning Mode**

   - Multi-audience summaries: Expert (technical), Student (with analogies), Policymaker (applications & risks).

   - Auto-generate flashcards and quizzes from each paper.

   - Personalized study roadmap guiding which papers to read first.

8. **Policy Brief Generator**

   - Summarize papers into policy-ready reports including applications, risks, and ethical issues.

   - Useful for policymakers, NGOs, or funding agencies.

9. **Visual Graph Exploration**

   - Interactive knowledge graphs connecting methods, datasets, authors, and institutions.

   - Click nodes to explore linked papers or collaborators.

   - Filter by year, domain, or author using D3.js or Neo4j Bloom.

10. **Cross-Domain Analogies**

   o   Generate analogies to make complex research concepts intuitive.

   o   Example: "Attention in NLP is like triaging patients in an ER."

11. **Personal Research Workspace**

   o   Save papers, notes, and queries.

   o   Auto-generate weekly reading summaries.

   o   Export structured literature reviews to PDF, Word, or LaTeX.

12. **Multi-Source Integration**

   o   Ingest PDFs and APIs (arXiv, PubMed, IEEE).

   o   Optional: integrate patent databases to connect research with industry innovations.

Tech Stack

- **Backend:** Python (FastAPI/Flask)

- **NLP/LLM:** HuggingFace Transformers (BART, LongT5, LLaMA-2), spaCy for NER

- **Graph DB:** Neo4j for graph storage & queries

- **Vector Store:** FAISS/Weaviate for semantic search

- **Frontend:** React + D3.js (interactive visualizations & dashboards)

- **Database:** PostgreSQL for user data and saved papers

- **Deployment:** Docker + AWS/GCP

Deliverables / Demo Flow

1. Upload papers → automatic entity extraction and processing.

2. Visualize knowledge graph → explore methods, datasets, authors, and institutions.

3. Ask research questions → receive evidence-backed answers.

4. Generate multi-audience summaries → expert, student, policymaker.

5. Access analytics dashboards → trending methods, collaboration networks, and forecasts.

6. Activate learning mode → view concept maps, flashcards, and quizzes.

7. Produce policy briefs → structured reports for decision-makers.

8. Save and export → personalized research workspace with literature review export.

Why This is Unique from Perplexity

1. **Purpose-Built for Research:** Dedicated to analyzing research papers with structured knowledge graphs (methods, datasets, authors, institutions).

2. **Evidence + Analytics:** Unlike Perplexity, answers come with direct citations and analytics like dataset frequencies, co-occurrence patterns, and collaboration networks.

3. **Knowledge Graph Exploration:** Interactive exploration from Dataset → Method → Paper → Author → Institution.

4. **Multi-Audience Summaries:** Built-in modes for Experts, Students, and Policymakers.

5. **Trend Forecasting & Collaboration Discovery:** Predicts emerging methods and suggests new research collaborations, not just descriptive answers.

Mandatory Requirements:

1. **Data Access:** Ensure that all research papers used are from **free and open-access sources** (arXiv, PubMed Central). Avoid paywalled publishers like IEEE or ACM unless you use their open-access subsets.

2. **Data Design:** Define a clear **database schema** to store metadata such as papers, authors, institutions, methods, and datasets. You might consider a **graph database**

(Neo4j) for representing relationships and an **SQL database** for structured metadata. Include an **ER or graph schema diagram** in your documentation.

3. **Data Preprocessing and Cleaning:** Demonstrate your pipeline for **PDF parsing, text extraction, and tokenization**, as well as how you handle noise (references, equations, figures). Treat this as part of your data management demonstration.

4. **Advanced SQL Reporting:** In your final submission, include at least **8 analytical reports or queries** demonstrating your ability to handle structured and relational data. For example:

    o   Top co-authorship pairs by publication count

    o   Most cited or trending research topics over time

    o   Paper count per institution or country

    o   Growth rate of research in specific AI subfields

    o   Top authors by publication frequency

    o   Datasets most frequently reused across papers

    o   Topic similarity or collaboration network density

    o   Recent methods emerging in multiple domains