Project Scope: Personal AI Research Assistant with Advanced Search and Literature Analysis

This scope integrates advanced search capabilities and the ability to analyze large volumes of research papers (100+ documents) for effective literature reviews while maintaining all core functionalities tailored to your needs.

Project Objective

To develop a scalable Personal AI Research Assistant that can:

1. Manage and analyze large datasets of academic materials.
2. Perform advanced mathematical and statistical modeling.
3. Automate workflows to reduce repetitive tasks.
4. Provide intelligent conversational assistance powered by GPT.
5. Deliver an OCD-friendly and user-friendly interface with modularity for future expansions.

Regrouped Functionalities

1. Document and Data Management

- **Purpose:** Centralize and organize research materials while supporting detailed search and analysis for effective literature reviews.
- **Key Features:**
    - **Advanced Search:**
        - Full-text search across research papers and documents.
        - Boolean operators (AND, OR, NOT) for refined queries.
        - Filtering by metadata (e.g., author, publication year, tags).
        - Fuzzy matching for typo tolerance.
    - **File Upload and Organization:**
        - Support manual and batch uploads of various document types (PDFs, Word, Excel, images).
        - Tagging and grouping for categorization.
    - **Cloud Integration:**
        - Manual syncing with iCloud, Google Drive, or Dropbox.
    - **Dataset Support:**
        - Analyze and process 100+ research papers in a single batch for literature reviews.
    - **Development Tools:**
        - ElasticSearch for full-text and metadata search.
        - Python libraries (PyPDF2, pdfplumber) for extracting text from PDFs.

2. Literature Review and Document Analysis

- **Purpose:** Facilitate literature reviews by extracting insights and summarizing large volumes of research papers.
- **Key Features:**
    - **Text Summarization:**
        - Summarize documents to highlight key findings, methodologies, and conclusions.
    - **Thematic Analysis:**
        - Extract recurring themes, keywords, and concepts across multiple papers.
    - **Citation Mapping:**

        - Identify relationships between papers, such as shared references or citations.
    - **Plagiarism Detection:**
        - Check originality by comparing text against existing research.
    - **Development Tools:**
        - NLP libraries (spaCy, NLTK) for text processing and thematic analysis.
        - OpenAI API for summarization and semantic understanding.

3. Advanced Physics, Mathematics, and Statistics

- **Purpose:** Provide analytical tools for validating research ideas and exploring hypotheses.
- **Key Features:**
    - **Mathematical Models:**
        - Tools for solving differential equations, optimization problems, and proofs.
    - **Statistical Analysis:**
        - Time-series analysis, regression, hypothesis testing, and multivariate analysis.
    - **Visualization:**
        - Generate plots, graphs, and charts for research output.
    - **Development Tools:**
        - Python libraries: NumPy, SciPy, Statsmodels, Matplotlib, and Plotly.

4. Workflow Automation

- **Purpose:** Automate repetitive tasks and streamline research workflows.
- **Key Features:**
    - **Auto-Tagging:**
        - Automatically tag and categorize files using AI-based rules.
    - **Report Generation:**
        - Create structured PDF or HTML reports summarizing literature review findings.
    - **Scheduled Backups:**
        - Automate local or cloud backups of files and datasets.
    - **Development Tools:**
        - Python scripting for automation.
        - ReportLab or WeasyPrint for report generation.

5. Conversational AI

- **Purpose:** Provide interactive, intelligent assistance for research and writing tasks.
- **Key Features:**
    - **GPT-Powered Conversational Assistant:**
        - Answer queries, validate hypotheses, and provide feedback on ideas.
    - **Writing Assistance:**
        - Grammar and style suggestions, referencing support (APA, MLA), and proofreading.
    - **External Data Integration:**
        - Query datasets and APIs from platforms like UN, IMF, and NASA.
    - **Development Tools:**
        - OpenAI API for conversational AI.
        - Grammarly or LanguageTool APIs for grammar checking.

6. Self-Learning and Adaptive Profiling

- **Purpose:** Enable the system to learn user preferences and improve over time.

- **Key Features:**
    - **Profile Building:**
        - Track preferred topics, frequent queries, and writing styles.
    - **Adaptive Suggestions:**
        - Recommend relevant research papers, tools, or topics.
    - **Proactive Alerts:**
        - Notify about new articles, datasets, or insights matching your research focus.
    - **Development Tools:**
        - SQLite/PostgreSQL for user profiling and lightweight memory.

7. Easy-to-Use Interface

- **Purpose:** Ensure a visually clean and user-friendly design tailored to your preferences.
- **Key Features:**
    - Customizable dashboard for quick access to core features.
    - Drag-and-drop functionality for file uploads and organization.
    - Minimalist, OCD-friendly interface with customizable themes (dark mode, fonts).
    - **Development Tools:**
        - Dash or Streamlit for frontend development.

Revised Architecture and Tools

Frontend:

- **Framework:** Dash or Streamlit.
- **Features:** File uploads, search, interactive dashboards, and conversational AI.

Backend:

- **Language:** Python.
- **Tools:**
    - File management: os, shutil.
    - Text analysis: PyPDF2, pdfplumber, spaCy.
    - Search: ElasticSearch.
    - AI Integration: OpenAI API.

Database:

- SQLite for metadata and user profiling.
- ElasticSearch for document indexing and advanced search.

APIs:

- OpenAI API for GPT-based conversational AI and summarization.
- APIs from UN, IMF, NASA for querying external datasets.

Updated Project Roadmap

Phase	Tasks	Duration
Phase 1: Document Management	Build document upload, tagging, and advanced search.	3 Weeks
Phase 2: Literature Analysis	Implement summarization, thematic analysis, and citation mapping.	3 Weeks
Phase 3: Math & Stats Tools	Develop modeling and statistical analysis tools.	3 Weeks
Phase 4: Workflow Automation	Automate tagging, backups, and report generation.	2 Weeks
Phase 5: Conversational AI	Integrate GPT for queries and writing assistance.	2 Weeks
Testing & Finalization	Test and refine the system.	1 Week

Total Duration: 14 Weeks.