ScholarHub

AI-Powered Research Collaboration Platform

Progression Report

Submitted by:

Hassan Ahmed Algizani
Student ID: g202403940

Supervised by:

Dr. Irfan Ahmad

King Fahd University of Petroleum and Minerals (KFUPM)

December 2025

Contents

Abstract

ScholarHub is a comprehensive, AI-powered research paper management and collaborative writing platform designed to streamline the academic research workflow. The platform addresses key challenges faced by researchers, including literature discovery, collaborative paper writing, reference management, and AI-assisted content generation.

This report presents the design, implementation, and features of ScholarHub, a full-stack web application built using modern technologies including React, FastAPI, PostgreSQL with vector embeddings, and real-time collaboration via CRDT-based synchronization. The platform integrates with multiple academic databases (ArXiv, PubMed, Semantic Scholar, etc.) for automated paper discovery and leverages OpenAI's language models for intelligent writing assistance.

Key contributions include a project-centric architecture that organizes research activities, real-time collaborative editing with conflict resolution, AI-powered literature review generation, and a multi-source academic paper discovery system with duplicate detection and relevance ranking.

Keywords: Research Collaboration, AI Writing Assistant, Literature Discovery, Real-time Collaboration, CRDT, Academic Platform

1 Introduction

1.1 Background and Motivation

Academic research has become increasingly collaborative and data-intensive. Researchers face numerous challenges in managing their workflow, including:

- **Literature Overload:** The exponential growth of published research makes it difficult to discover and track relevant papers.
- **Collaboration Barriers:** Teams working across institutions need effective tools for real-time collaboration on manuscripts.
- **Reference Management:** Organizing citations and generating bibliographies remains tedious and error-prone.
- **Writing Efficiency:** Academic writing is time-consuming and could benefit from AI assistance while maintaining scholarly rigor.

While various tools address individual aspects of research workflow (e.g., Zotero for references, Overleaf for LaTeX collaboration, Google Scholar for discovery), there exists a need for an integrated platform that combines these capabilities with modern AI assistance.

1.2 Project Objectives

ScholarHub aims to provide a unified platform that:

1. Enables project-based organization of research activities
2. Supports real-time collaborative paper writing in LaTeX and rich-text formats
3. Automates literature discovery from multiple academic databases
4. Provides AI-powered writing assistance and literature review generation
5. Facilitates team communication through integrated discussion channels
6. Offers video conferencing with automatic transcription for research meetings

1.3 Report Organization

The remainder of this report is organized as follows: Section 2 reviews related work and existing solutions. Section 3 presents the system architecture and technology stack. Section 4 details the implementation of core features. Section 5 demonstrates the platform through screenshots. Section 6 discusses challenges and lessons learned. Section 7 concludes with future work directions.

2 Related Work and Background

2.1 Existing Research Tools

Several tools currently address aspects of the research workflow:

Reference Managers: Tools like Zotero, Mendeley, and EndNote help researchers organize papers and generate citations. However, they primarily focus on reference management without integrated writing or discovery capabilities.

Collaborative Writing: Overleaf provides excellent LaTeX collaboration but lacks AI assistance and integrated reference discovery. Google Docs offers real-time collaboration but isn't optimized for academic writing.

Literature Discovery: Google Scholar, Semantic Scholar, and Connected Papers help discover related research, but require manual effort to integrate findings into projects.

AI Writing Assistants: Tools like Grammarly and ChatGPT offer writing assistance, but aren't designed for the specific needs of academic research with proper citations and domain knowledge.

2.2 Technical Foundations

2.2.1 Real-time Collaboration

Conflict-free Replicated Data Types (CRDTs) enable real-time collaborative editing without central coordination. Yjs, a JavaScript CRDT implementation, allows multiple users to edit documents simultaneously with automatic conflict resolution.

2.2.2 Vector Embeddings for Search

Modern semantic search uses neural network embeddings to represent text as high-dimensional vectors. PostgreSQL's pgvector extension enables efficient similarity search on these embeddings, powering intelligent paper recommendations.

2.2.3 Large Language Models

OpenAI's GPT models provide natural language understanding and generation capabilities that enable intelligent writing assistance, query enhancement, and literature summarization.

3 System Architecture

3.1 Overview

ScholarHub follows a microservices-inspired architecture with clear separation between frontend, backend, and collaboration services. Figure 1 illustrates the high-level system architecture.

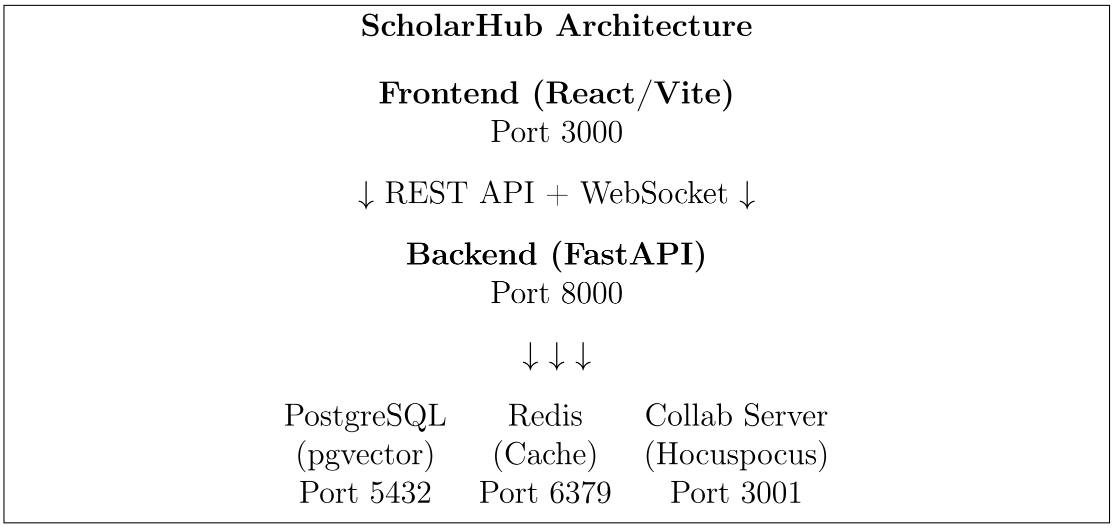


Figure 1: High-level System Architecture

3.2 Technology Stack

Table 1 summarizes the technologies used in each layer of the application.

Table 1: Technology Stack

Layer	**Technology**	**Purpose**
Frontend	React 18, TypeScript, Vite TailwindCSS TipTap / CodeMirror React Query	User interface Styling Rich-text and LaTeX editors Server state management
Backend	FastAPI (Python) SQLAlchemy 2.0 Pydantic OpenAI SDK	REST API framework ORM for database access Data validation AI integration
Database	PostgreSQL 15 pgvector Redis 7	Primary data store Vector similarity search Caching and sessions
Collaboration	Hocuspocus Yjs	Real-time sync server CRDT implementation
Infrastructure	Docker Compose Daily.co	Container orchestration Video conferencing

3.3 Database Schema

The database consists of 26 models organized into logical groups:

Core Entities:

- `User` – User accounts with authentication
- `Project` – Research projects as organizational units
- `ResearchPaper` – Papers within projects
- `Document` – Uploaded PDFs and documents
- `Reference` – Academic paper references

Collaboration:

- `ProjectMember` – Project membership with roles
- `Comment` – Threaded comments on content
- `CollaborationSession` – Real-time editing sessions

AI and Search:

- `DocumentChunk` – Chunked content with vector embeddings
- `AIChatSession` – AI conversation history
- `AIArtifact` – Generated content (summaries, reviews)

4 Implementation and Features

4.1 Project Management

ScholarHub uses a project-centric architecture where all research activities are organized within projects. Each project contains:

- Research papers (LaTeX or rich-text format)
- Reference library with approved citations
- Team members with role-based permissions (owner, editor, viewer)
- Discussion channels for team communication
- Sync space for video meetings

Projects support keywords and scope definitions that inform AI recommendations and literature discovery.

4.2 Collaborative Paper Editor

The platform supports two editor modes:

LaTeX Editor: Built on CodeMirror 6 with syntax highlighting, auto-completion, and real-time PDF preview. Supports collaborative editing via Yjs CRDT synchronization.

Rich-Text Editor: Built on TipTap (ProseMirror) with formatting tools, tables, images, and math equations via KaTeX. Also supports real-time collaboration.

Real-time synchronization is handled by a Hocuspocus server that:

1. Maintains document state using Yjs CRDTs
2. Broadcasts changes to all connected clients
3. Persists state to Redis for recovery
4. Handles conflict resolution automatically

4.3 Literature Discovery

The discovery system aggregates results from multiple academic APIs:

- ArXiv (preprints in physics, math, CS)
- PubMed (biomedical literature)
- Semantic Scholar (AI-powered academic search)
- OpenAlex (open scholarly metadata)
- CrossRef (DOI resolution and metadata)
- ScienceDirect (Elsevier publications)

Key features include:

- Smart query enhancement using GPT for better search terms
- Duplicate detection across sources
- Relevance ranking based on project keywords
- PDF access detection via Unpaywall
- Configurable source preferences per project

4.4 AI Integration

The platform integrates OpenAI's language models for several features:

Writing Assistant: Provides context-aware suggestions, helps rephrase text, and generates content based on project references.

Literature Review Generation: Synthesizes selected references into coherent literature review sections with proper citations.

Chat Interface: Answers questions about project content, explains concepts, and assists with research tasks.

Query Enhancement: Improves search queries for better literature discovery results.

Meeting Transcription: Uses Whisper API to transcribe recorded meetings and extract action items.

4.5 Reference Management

References are managed at both project and paper levels:

- Automatic metadata extraction from DOIs
- Citation parsing from various formats (BibTeX, RIS, etc.)
- Project reference library with approval workflow
- Paper-specific reference linking
- Bibliography generation in multiple styles

4.6 Team Collaboration

Beyond real-time editing, the platform supports:

- **Discussion Channels:** Organized conversation threads with file sharing
- **Comments:** Inline and section-level comments on papers
- **Notifications:** Activity feed for project updates
- **Sync Space:** Video meetings via Daily.co with recording and transcription

5 System Demonstration

This section presents screenshots of the ScholarHub platform demonstrating key features.

5.1 Authentication

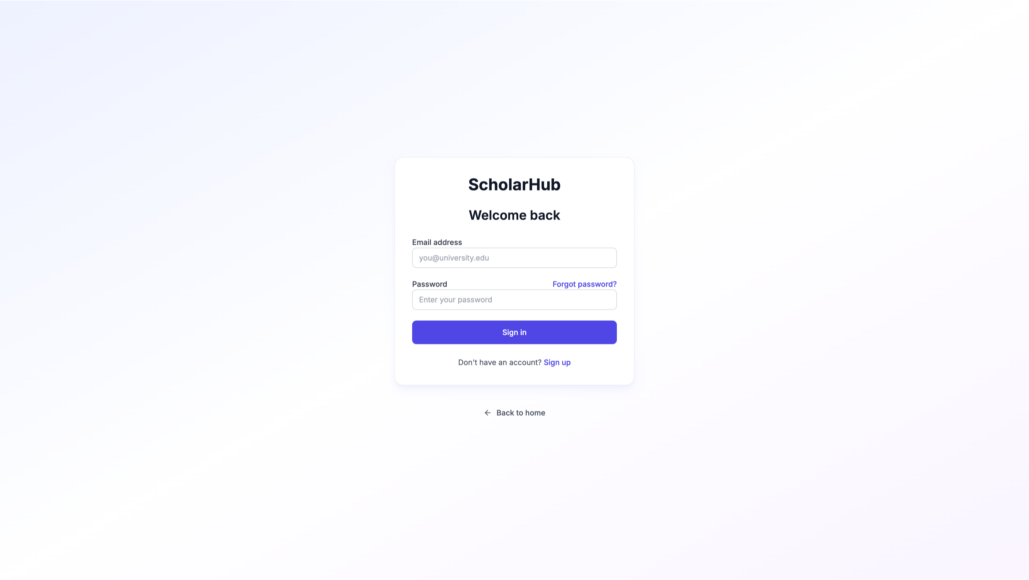
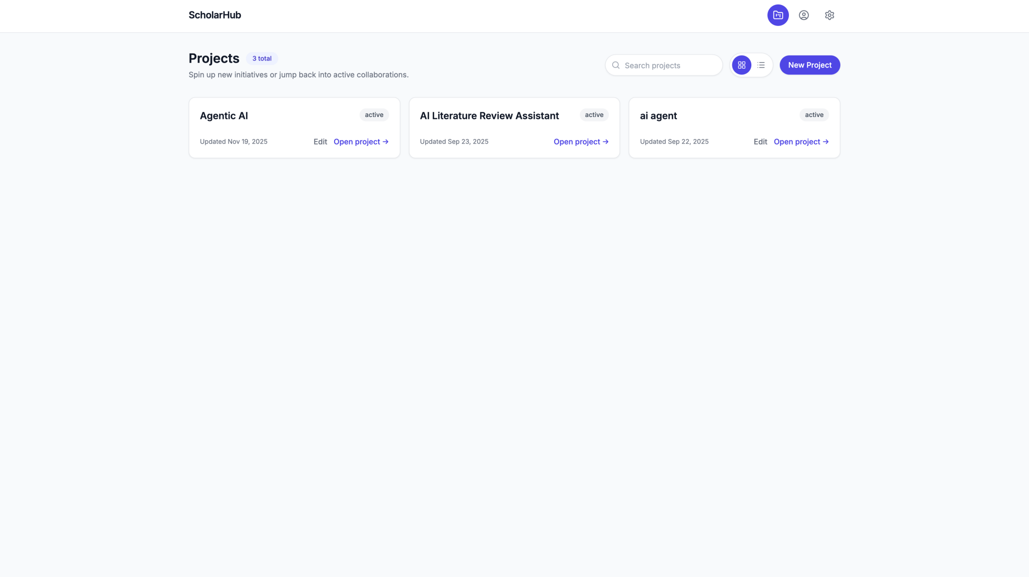


Figure 2: Login Page – Secure authentication with email and password

The login page provides a clean, modern interface for user authentication. Users can register new accounts or sign in with existing credentials.

5.2 Projects Dashboard



Projects 3 total

Spin up new initiatives or jump back into active collaborations.

Agentic AI active
Updated Nov 19, 2025 Edit Open project →

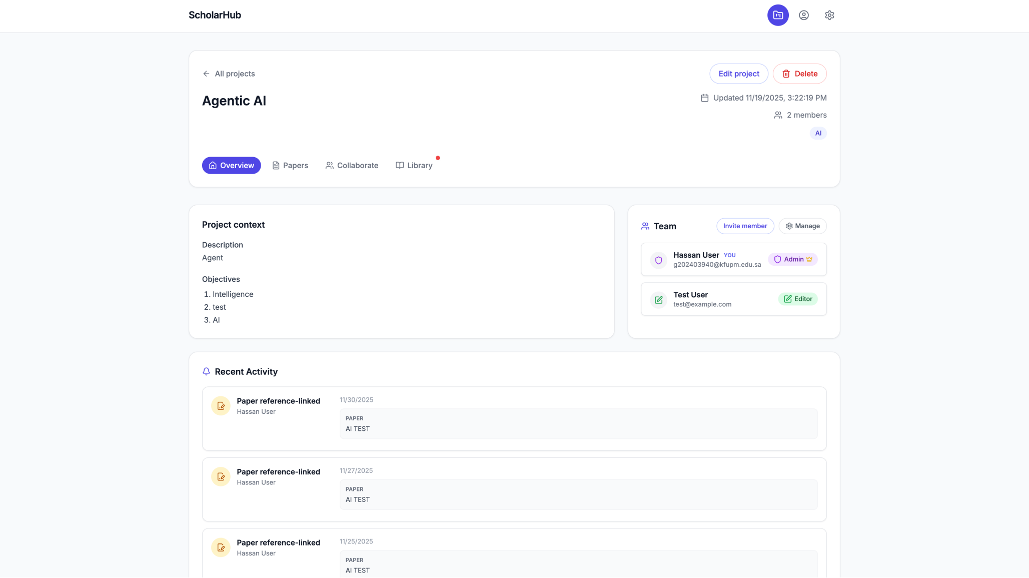
AI Literature Review Assistant active
Updated Sep 23, 2025 Open project →

ai agent active
Updated Sep 22, 2025 Edit Open project →

Figure 3: Projects Dashboard – Overview of all research projects

The dashboard displays all projects accessible to the user, showing project titles, descriptions, and quick access to recent activity.

5.3 Project Overview



Agentic AI

Updated 11/19/2025, 3:22:19 PM
2 members
AI

Project context

Description
Agent

Objectives
1. Intelligence
2. test
3. AI

Team

Hassan User YOU Admin
g202403940@kfupm.edu.sa

Test User Editor
test@example.com

Recent Activity

Paper reference-linked
Hassan User
11/30/2025
PAPER
AI TEST

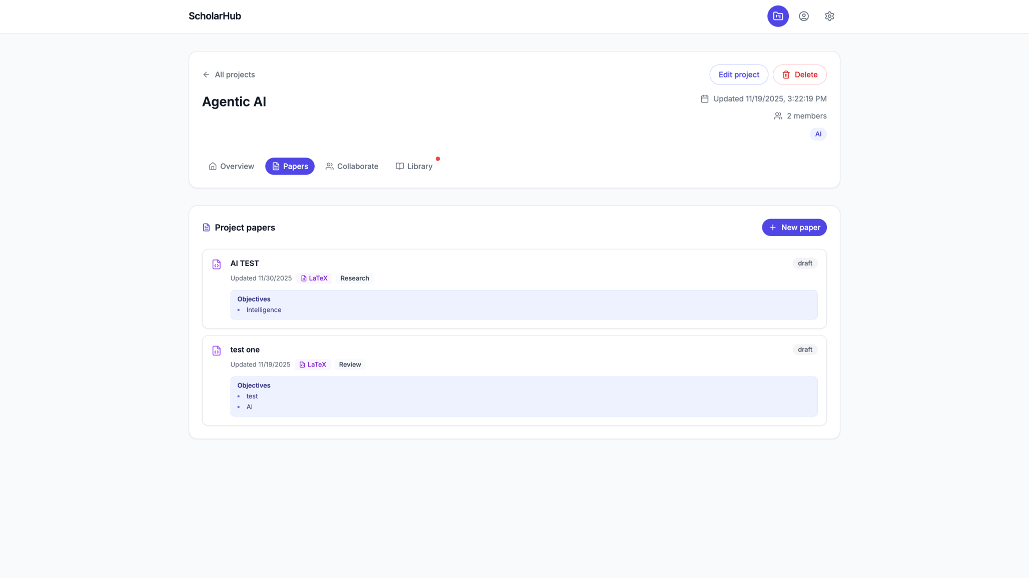
Paper reference-linked
Hassan User
11/27/2025
PAPER
AI TEST

Paper reference-linked
Hassan User
11/25/2025
PAPER
AI TEST

Figure 4: Project Overview – Project details with team members and recent activity

Each project has an overview page showing project description, keywords, team members with roles, and recent activity feed.

5.4 Papers Management



ScholarHub

← All projects

Agentic AI

Edit project

Delete

Updated 11/19/2025, 3:22:19 PM

2 members

AI

Overview

Papers

Collaborate

Library

Project papers

+ New paper

AI TEST

draft

Updated 11/30/2025 LaTeX Research

Objectives

- Intelligence

test one

draft

Updated 11/19/2025 LaTeX Review

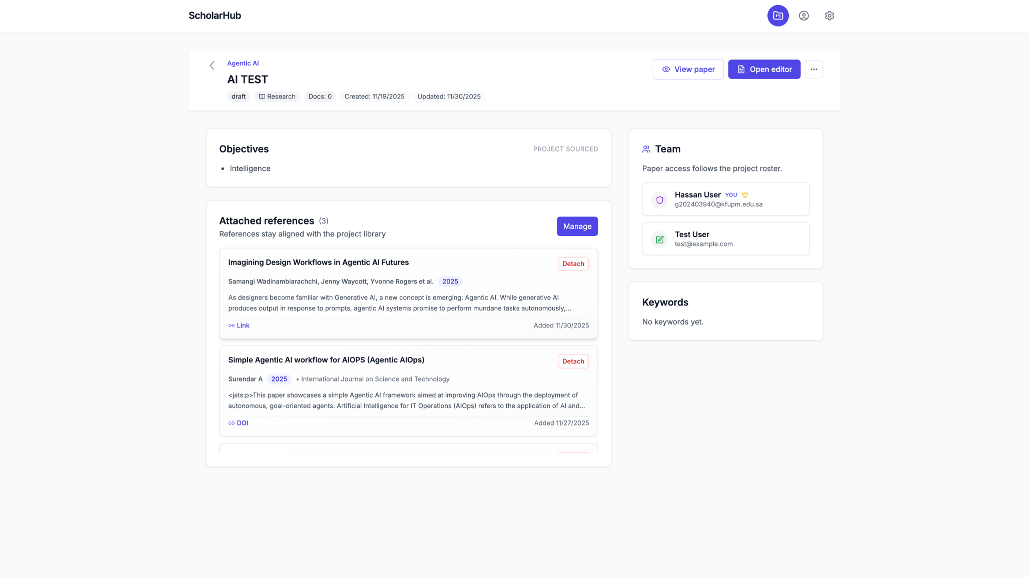
Objectives

- test
- AI

Figure 5: Papers List – All papers within a project

The papers tab lists all research papers within a project, with options to create new papers using templates or from scratch.

5.5 Paper Editor



ScholarHub

Agentic AI

AI TEST

draft Research Docs: 0 Created: 11/19/2025 Updated: 11/30/2025

View paper

Open editor

···

Objectives

PROJECT SOURCED

- Intelligence

Attached references (3)

References stay aligned with the project library

Manage

Imagining Design Workflows in Agentic AI Futures

Detach

Samangi Wadinambiarachchi, Jenny Waycott, Yvonne Rogers et al. 2025

As designers become familiar with Generative AI, a new concept is emerging: Agentic AI. While generative AI produces output in response to prompts, agentic AI systems promise to perform mundane tasks autonomously,...

Link

Added 11/30/2025

Simple Agentic AI workflow for AIOPS (Agentic AIOps)

Detach

Surendar A 2025 • International Journal on Science and Technology

<jats:p>This paper showcases a simple Agentic AI framework aimed at improving AIOps through the deployment of autonomous, goal-oriented agents. Artificial Intelligence for IT Operations (AIOps) refers to the application of AI and...

DOI

Added 11/27/2025

Team

Paper access follows the project roster.

Hassan User YOU

g202403940@kfupm.edu.sa

Test User

test@example.com

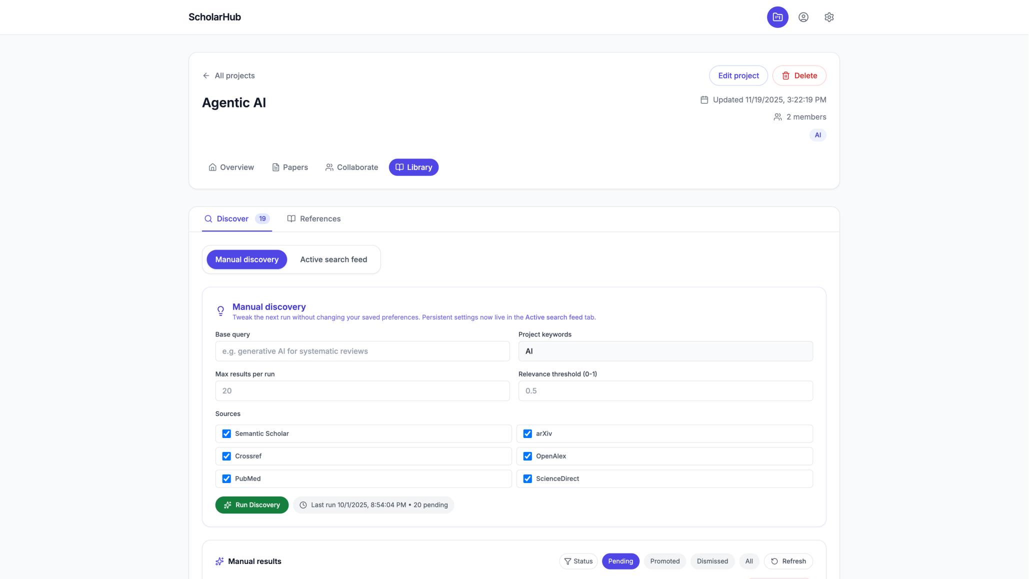
Keywords

No keywords yet.

Figure 6: Paper Editor – Collaborative writing with AI assistance

The paper editor supports both LaTeX and rich-text editing with real-time collaboration. The interface includes formatting tools, AI chat panel, and reference integration.

5.6 Research Library



Agentic AI

Updated 11/19/2025, 3:22:19 PM

2 members

AI

Manual discovery

Tweak the next run without changing your saved preferences. Persistent settings now live in the Active search feed tab.

Base query: e.g. generative AI for systematic reviews

Project keywords: AI

Max results per run: 20

Relevance threshold (0-1): 0.5

Sources

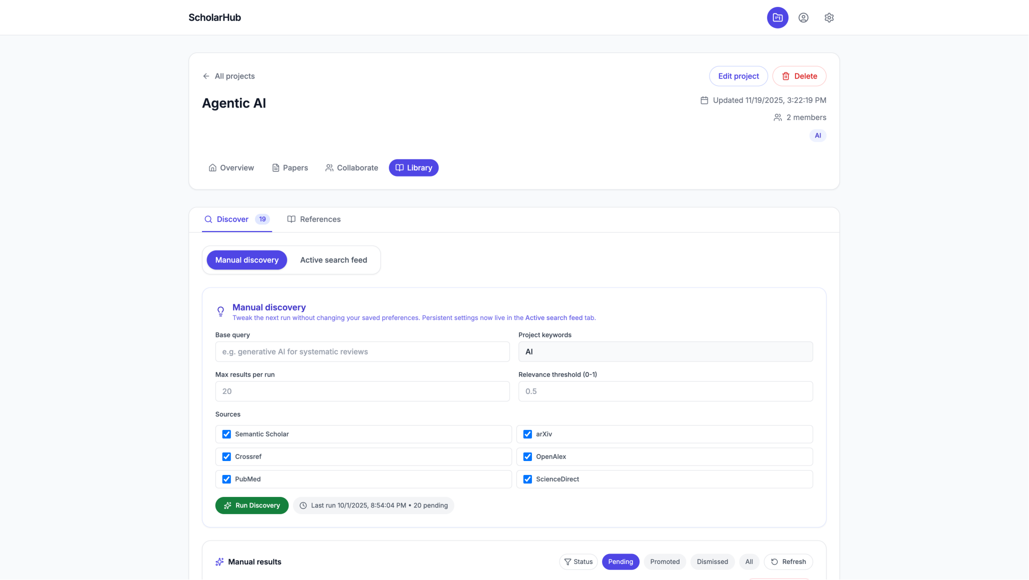
☑ Semantic Scholar
☑ arXiv
☑ Crossref
☑ OpenAlex
☑ PubMed
☑ ScienceDirect

Last run 10/1/2025, 8:54:04 PM • 20 pending

Figure 7: Research Library – Manage references and discover papers

The library provides a unified interface for managing references, viewing uploaded documents, and discovering new papers.

5.7 Paper Discovery



Agentic AI

Updated 11/19/2025, 3:22:19 PM

2 members

AI

Manual discovery

Tweak the next run without changing your saved preferences. Persistent settings now live in the Active search feed tab.

Base query: e.g. generative AI for systematic reviews

Project keywords: AI

Max results per run: 20

Relevance threshold (0-1): 0.5

Sources

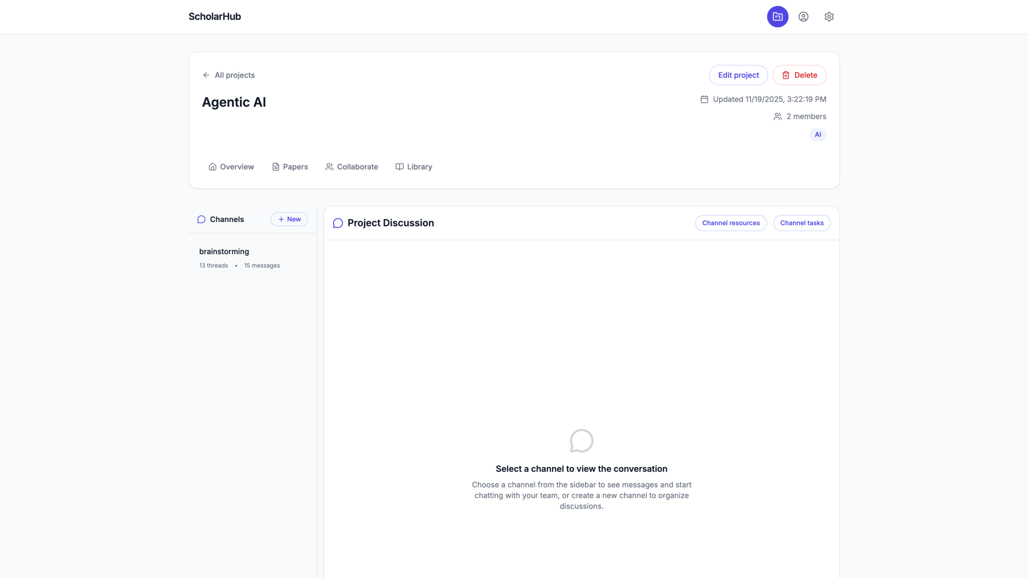
☑ Semantic Scholar
☑ arXiv
☑ Crossref
☑ OpenAlex
☑ PubMed
☑ ScienceDirect

Last run 10/1/2025, 8:54:04 PM • 20 pending

Figure 8: Paper Discovery – AI-powered literature search across multiple databases

The discovery feature searches multiple academic databases simultaneously, ranks results by relevance, and allows users to add papers directly to their project library.

5.8 Team Discussion



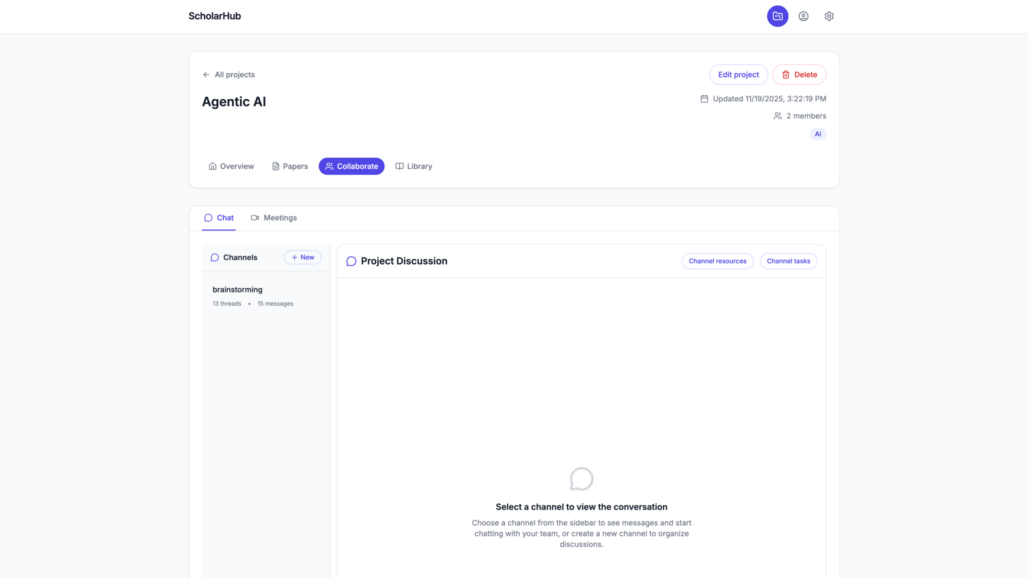
Select a channel to view the conversation

Choose a channel from the sidebar to see messages and start chatting with your team, or create a new channel to organize discussions.

Figure 9: Discussion Channels – Team communication with organized threads

Discussion channels enable team communication with support for file attachments, mentions, and organized conversation threads.

5.9 Collaboration and Sync Space



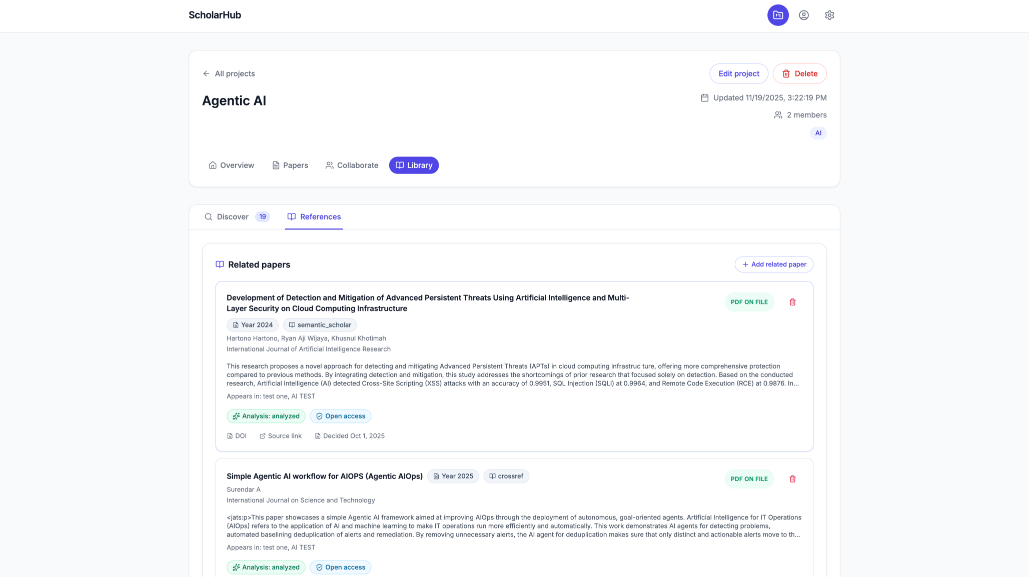
Select a channel to view the conversation

Choose a channel from the sidebar to see messages and start chatting with your team, or create a new channel to organize discussions.

Figure 10: Collaboration – Video meetings and real-time sync

The collaboration tab provides access to video meetings with automatic transcription and action item extraction.

5.10 Reference Management



Development of Detection and Mitigation of Advanced Persistent Threats Using Artificial Intelligence and Multi-Layer Security on Cloud Computing Infrastructure

Year 2024 · semantic_scholar · PDF ON FILE

Hartono Hartono, Ryan Aji Wijaya, Khusnul Khotimah

International Journal of Artificial Intelligence Research

This research proposes a novel approach for detecting and mitigating Advanced Persistent Threats (APTs) in cloud computing infrastruc ture, offering more comprehensive protection compared to previous methods. By integrating detection and mitigation, this study addresses the shortcomings of prior research that focused solely on detection. Based on the conducted research, Artificial Intelligence (AI) detected Cross-Site Scripting (XSS) attacks with an accuracy of 0.9951, SQL Injection (SQLI) at 0.9964, and Remote Code Execution (RCE) at 0.9876. In...

Appears in: test one, AI TEST

Analysis: analyzed · Open access

DOI · Source link · Decided Oct 1, 2025

Simple Agentic AI workflow for AIOPS (Agentic AIOps) · Year 2025 · crossref · PDF ON FILE

Surendar A

International Journal on Science and Technology

<jats:p>This paper showcases a simple Agentic AI framework aimed at improving AIOps through the deployment of autonomous, goal-oriented agents. Artificial Intelligence for IT Operations (AIOps) refers to the application of AI and machine learning to make IT operations run more efficiently and automatically. This work demonstrates AI agents for detecting problems, automated baselining deduplication of alerts and remediation. By removing unnecessary alerts, the AI agent for deduplication makes sure that only distinct and actionable alerts move to th...

Appears in: test one, AI TEST

Analysis: analyzed · Open access

Figure 11: References – Comprehensive reference management with metadata

The references view shows approved citations with full metadata, allowing easy management and export in various formats.

6 Challenges and Lessons Learned

6.1 Technical Challenges

Real-time Collaboration: Implementing conflict-free real-time editing required careful integration of Yjs CRDTs with both editor frameworks. Handling network disconnections and state recovery was particularly challenging.

Multi-source Discovery: Each academic API has different rate limits, response formats, and data quality. Building a unified interface required extensive normalization and error handling.

Vector Search Performance: As the document corpus grows, maintaining fast similarity search required careful indexing strategies with pgvector.

AI Context Management: Providing relevant context to AI models while staying within token limits required intelligent chunking and retrieval strategies.

6.2 Lessons Learned

1. **Start with Core Features:** Focus on essential functionality before adding advanced features.

2. **API Abstractions:** Abstract external APIs behind unified interfaces to simplify maintenance.

3. **Type Safety:** TypeScript on the frontend and Pydantic on the backend caught many errors early.

4. **Feature Flags:** Implementing feature flags allowed gradual rollout of new capabilities.

7 Conclusion and Future Work

7.1 Summary

This report presented ScholarHub, an AI-powered research collaboration platform that integrates paper writing, literature discovery, reference management, and team collaboration into a unified experience. The platform demonstrates the potential of combining modern web technologies with AI capabilities to enhance research productivity.

Key achievements include:

- Project-centric architecture for organized research
- Real-time collaborative editing using CRDTs
- Multi-source literature discovery with intelligent ranking
- AI-powered writing assistance and literature review generation
- Integrated team communication and video meetings

7.2 Future Work

Several directions for future development include:

1. **Citation Network Analysis:** Visualize relationships between papers and identify influential works.
2. **Automated Peer Review:** AI-assisted feedback on manuscript quality and completeness.
3. **Mobile Application:** Native mobile apps for reading and annotation on the go.
4. **Institutional Integration:** SSO and library access integration for universities.
5. **Offline Support:** Enable offline editing with automatic synchronization.
6. **Advanced Analytics:** Track reading progress, writing statistics, and team productivity.

7.3 Conclusion

ScholarHub represents a comprehensive solution to the challenges of modern academic research. By combining collaborative writing, AI assistance, and literature management in a single platform, it has the potential to significantly improve research workflow efficiency. The modular architecture and modern technology stack provide a solid foundation for continued development and enhancement.

References

1. Kleppmann, M., & Beresford, A. R. (2017). A conflict-free replicated JSON datatype. *IEEE Transactions on Parallel and Distributed Systems*, 28(10), 2733-2746.

2. Brown, T., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.

3. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *arXiv preprint arXiv:1908.10084*.

4. FastAPI Documentation. (2024). FastAPI framework, high performance, easy to learn. https://fastapi.tiangolo.com/

5. React Documentation. (2024). React – A JavaScript library for building user interfaces. https://react.dev/

6. Yjs Documentation. (2024). Shared data types for building collaborative software. https://yjs.dev/

7. OpenAI. (2024). GPT-4 Technical Report. https://openai.com/research/gpt-4

8. pgvector. (2024). Open-source vector similarity search for Postgres. https://github.com/pgvector/pgvector