A Hybrid RAG System with Graph Based Context Retrieval

Our Project is the Graph RAG Powered Academic Assistant, an AI based tool that augments Retrieval Augmented Generation (RAG) with citation graph analysis. It ingests research papers, builds citation aware networks, and provides natural language answers with traceable citations. Key features include dual retrieval (semantic + citation graph), citation trail visualization, and explainability modules.

Pant Divyansh

Singh Gagan Preet

Tiwari Sourav

Valani Rameshkumar Premji

POTENTIAL ADVISORS: Dr. Xiao Hu

Date: September 27, 2025

|  |  |  |  |
| --- | --- | --- | --- |
| Revision History Table | | Template Date 6/15/2022 | |
| Version | Summary of Changes | | Date |
| 0.1 | Draft Introduction added | | 09/20/25 |
| 0.3 | Template info removed; first draft of all sections completed | | 09/22/25 |
| 0.9 | All sections updated during team review | | 09/26/25 |
| 1.0 | Version submitted for signatures | | 09/27/25 |

Contents

[1. Executive Summary 3](#_Toc2002362744)

[2. User/Market research 4](#_Toc770924471)

[3. Product Features 5](#_Toc1636394344)

[Feature 1: PDF Parsing & Chunking 6](#_Toc757297297)

[Feature 2: Citation Graph Builder 6](#_Toc1501082868)

[Feature 3: Dual Retrieval (Semantic + Citation Graph) 6](#_Toc1432808667)

[Feature 4: Explainability, Monitoring & Tracing Module 6](#_Toc1446445651)

[Feature 5: Web UI 6](#_Toc1500406552)

[Stretch Feature (Optional): If time permits 6](#_Toc1660190463)

[4. Project Timeline & Gannt Chart 7](#_Toc1942912498)

[5. Ethics 10](#_Toc383641654)

[6. Approvals 12](#_Toc1685499613)

[7. Appendix 13](#_Toc1303960465)

[A. Advisor Engagement 14](#_Toc1448849615)

[1) Project Team Responsibilities 14](#_Toc783334649)

[2) Faculty Advisor Responsibilities 14](#_Toc37966632)

[B. Ground Rules 14](#_Toc1794592270)

# Executive Summary

*The Executive Summary was written by Valani Rameshkumar.*

Our product is a Graph RAG Powered Academic Assistant that transforms how students and researchers engage with scholarly literature. It combines semantic retrieval and citation graph expansion to deliver explainable, transparent, and source aware answers to user queries. The system parses uploaded PDFs, extracts and chunks content, and builds a citation graph that represents relationships among papers. Key features include dual retrieval (semantic similarity plus citation graph traversal), explainability modules that trace citation paths, confidence scoring, and an interactive user interface where users can view retrieved passages, graph visualizations, and the reasoning behind each answer. In short, the assistant helps transform dense research papers into clear, verifiable insights.

This product is needed because existing Retrieval-Augmented Generation (RAG) systems often produce hallucinated or unverifiable claims, undermining trust in AI research tools. According to Ji et al. (2023), hallucinations in natural language generation are among the top barriers to adoption in academia and law, where source transparency is essential. Our system addresses this gap by integrating citation graphs, ensuring that every answer is supported by verifiable references. Unlike conventional systems that return isolated text chunks, our assistant explains how answers are derived by showing citation trails. This uniquely positions our solution to benefit not only students but also the broader academic community by improving trustworthiness, transparency, and efficiency in navigating research literature.

The project will be developed in seven structured phases, ranging from literature review and environment setup to retrieval system design, LLM integration, and UI development. Work will take place collaboratively, with each team member responsible for distinct modules. At the end of the semester, we will deliver a fully functional prototype capable of ingesting PDFs, retrieving contextual answers, and displaying citation-backed explanations. Table 1 shows the preliminary division of responsibilities. Notably, the parsing, explainability, and user interface modules have been distributed across multiple members to balance technical precision with usability.

|  |  |
| --- | --- |
| Team Member | Feature responsibility |
| Valani Rameshkumar | Backend/Infrastructure, embeddings, evaluation |
| Singh Gagan Preet | Citation graph builder, traversal, fusion, tuning |
| Tiwari Sourav | PDF parsing, visualization, explainability |
| Pant Divyansh | LLM integration, UI, documentation |

Table 1 Preliminary Subsystem Responsibilities

# User/Market research

*This section was written by Singh Gagan.*

Navigating research literature is a common challenge faced by students and early-stage researchers. Each semester, students are expected to review dozens of technical papers, but existing tools like keyword search or generic academic databases often fail to provide context on how ideas evolve across citations. This creates barriers to comprehension, slows down literature reviews, and reduces the quality of academic writing.

Existing Academic Tools:

* Google Scholar provides access to citations but lacks the ability to answer natural language questions or explain reasoning paths.
* Vanilla RAG based prototypes in academic settings improve retrieval but suffer from hallucinations and lack transparency in reasoning.

Our project is unique for an academic project because it integrates both semantic search and citation graph analysis, ensuring that students can not only retrieve relevant passages but also understand how one paper builds on another. This dual retrieval approach supports learning explainability, which is essential in academic research training.

User Insights:

* Students spend significant time tracing citations manually to understand the progression of research ideas.
* There is low trust in AI answers when sources and reasoning are not clearly visible.
* Both groups prefer tools that support learning rather than replacing critical thinking.

How Our Project Addresses These Needs:

* By displaying linked citation paths (e.g., Vaswani → Bahdanau → Luong), students can visualize the academic lineage of ideas.
* Every AI generated answer is paired with source attribution and chunk level tagging, enabling students to verify claims.
* The citation graph visualization provides a new way to explore academic literature, turning static reading into an interactive process.

It provides students and researchers with a guided, explainable assistant that improves comprehension of scholarly literature while maintaining academic rigor and integrity.

# Product Features

*This section was written by Tiwari Sourav.*

The Graph RAG Powered Academic Assistant is designed to support students and researchers in navigating academic literature by combining semantic search with citation graph analysis. The product emphasizes explainability, traceability, and academic integrity. Below are the key features that define the scope of the project.

## Feature 1: PDF Parsing & Chunking

The system will accept academic papers in PDF format and automatically parse them into semantically meaningful text chunks with metadata such as authors, year, and references. This ensures that each piece of content retains context while being optimized for retrieval. By structuring papers into smaller, annotated segments, the assistant can provide more accurate and focused answers to user queries.

## Feature 2: Citation Graph Builder

A citation graph will be constructed using external APIs such as OpenAlex, where nodes represent papers and edges represent citation relationships. This feature allows users to see how ideas are connected across papers, tracing the flow of knowledge through multiple generations of research. The interactive visualization of this graph will make it easier for students and researchers to follow academic lineages.

## Feature 3: Dual Retrieval (Semantic + Citation Graph)

Unlike traditional RAG systems that rely only on semantic similarity, our assistant combines vector based semantic retrieval with citation graph traversal. This dual approach ensures both contextual relevance and citation supported accuracy. The fusion of results reduces hallucinations while improving trust in AI generated answers by grounding them in verifiable sources.

## Feature 4: Explainability, Monitoring & Tracing Module

This module enhances transparency by showing users how the system arrived at an answer. Features include citation trails (e.g., Vaswani → Bahdanau → Luong), chunk-level attribution, and confidence scoring. By making reasoning paths explicit, students can verify claims directly in the source material, ensuring academic integrity and building trust in the assistant’s outputs.  
LLM Request Tracing: Phoenix captures every LLM call along with inputs, outputs, and execution metadata.

## Feature 5: Web UI

The project includes a simple, interactive UI web interface. Users will be able to upload research papers, ask natural language questions, and view answers paired with citations and graphs. The UI is designed for accessibility, making it suitable for academic settings such as classrooms, labs, and independent study.

## Stretch Feature (Optional): If time permits

Additional enhancements may include cloud deployment on AWS for scalability, richer visualization tools such as confidence histograms, and modular multi-agent extensions for specialized tasks like parsing, retrieval, and reasoning. These will be considered if the core features are stable and functional before the final milestone.

* Cloud deployment on AWS (S3, Lambda, EC2) for scalability.
* Advanced visualizations (confidence score histograms, chunk length distribution).
* Multi agent extension for modular workflows (Document Agent, Citation Agent, Retrieval Agent, etc.).

# Project Timeline & Gannt Chart

*This section was written by Pant Divyansh.*

The project will progress in seven phases:

* Week 1–2: Literature review, environment setup, repository initialization.
* Week 2–3: PDF parsing & citation graph builder.
* Week 3–4: Dual retrieval engine development.
* Week 4–5: LLM integration and explainability module.
* Week 5–6: UI development.
* Week 6–7: Evaluation, hyperparameter tuning.
* Week 7–9: Cloud deployment, testing, documentation, final demo.

|  |  |
| --- | --- |
| Milestone | Date |
|  |  |
| Team Formation | 15/09/2025 |
| Project Proposal Finalization | 1/10/2025 |
| Signed Proposal | 1/10/2025 |
| Foundations Stage | 10/10/2025 |
| * Task 1.1 - Literature Review & Design |  |
| * Task 1.2 - Environment & Repository Setup |  |
| Data Layer | 10/17/2025 |
| * Task 2.1 - PDF Parsing & Chunking |  |
| * Task 2.2 - Citation Graph Builder |  |
| Retrieval Layer | 10/30/2025 |
| * Task 3.1 - Vector Embeddings & Storage |  |
| * Task 3.2 - Graph Traversal & Fusion Engine |  |
| Reasoning & Explainability | 11/07/2025 |
| * Task 4.1 - LLM Response Integration |  |
| * Task 4.2 - Explainability & Tracing Module |  |
| Interface & Demo | 11/14/2025 |
| * Task 5.1 – User Interface (UI) |  |
| * Task 5.2 - Visualization Enhancements |  |
| Evaluation & Optimization | 11/21/2025 |
| * Task 6.1 - Metrics Evaluation |  |
| * Task 6.2 - Hyperparameter Tuning |  |
| Wrap-up & Final Deliverables | 11/28/2025 |
| * Task 7.1 - AWS Setup & Storage |  |
| * Task 7.2 - Lambda Wrapper for PDF Ingestion |  |
| * Task 7.3 - Deploy UI via Docker on EC2 |  |
| * Task 7.4 - Testing & Demo Prep |  |
| Documentations | 11/28/2025 |
| iShowcase | 12/05/2025 |

Table 3: Milestone Schedule

A screenshot of a computer screen

AI-generated content may be incorrect.

# Ethics

*This section was written by All Team Members.*

We recognize potential ethical risks such as data misuse, bias in retrieval, and privacy concerns with academic content. Our system does not store sensitive personal data. We mitigate risks through citation traceability, transparent attribution, and compliance with academic integrity. We also conducted a data ethics questionnaire (see Appendix) to assess potential misuse, including risks of misinformation and biased predictions.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Question** | **Generally** | **Data Breach** |
| 1 | Could a user sell drugs or other illegal items on your platform? | N | N |
| 2 | Could a user of your platform engage in sex trafficking? | N | N |
| 3 | Could a user sell class notes or cheat on their homework on your platform? | N | N |
| 4 | Could a stalker use your project to find someone? | N | N |
| 5 | Could your app be used to spy on or track individuals? | N | N |
| 6 | Could your app/software access the camera or microphone and record things without users being aware? | N | N |
| 7 | If someone uses your platform, could they be re-traumatized or have their mental health impacted in some way? | N | N |
| 8 | Could your algorithm promote material that would traumatize or upset individuals? | N | N |
| 9 | Would your users be upset if the data you collect was given to someone else? | N | N |
| 10 | Could a data leak potentially lead to identity theft? | N | N |
| 11 | If your site was hacked, would users of that product potentially lose their job, spouse, or family? | N | N |
| 12 | Should there be an age limitation on your product? | N | N |
| 13 | Could someone use your product to find, contact, and potentially commit elder abuse? | N | N |
| 14 | If the data on your platform was breached, could it be used to blackmail the users? | N | N |
| 15 | Does the existence of your project imply that a particular racial group, gender, religion or other protected category is inherently bad, gross, or unwanted? | N | N |
| 16 | Could your product be used to commit hate crimes against a specific group? | N | N |
| 17 | Does the primary content of your game or algorithm focus on something considered deeply unethical? | N | N |
| 18 | Does your game or software contain race, gender, or other stereotypes? | N | N |
| 19 | Could users of your app scam other individuals? | N | N |
| 20 | Is your particular algorithm biased towards predicting correctly only for one race, gender, or other group? | N | N |
| 21 | Are the users of your project, players of your game, or those being surveyed for your data aware of how their data will be used? | N | N |
| 22 | What are the possible misinterpretations of your results? For example - would a white supremacist or misogynist be stoked about your results if they misinterpreted it? | N | N |
| 23 | Does the use or purchase of your data potentially contribute to a dangerous group or regime? | N | N |
| 24 | Could your virtual reality environment cause injury to the user? | N | N |
| 25 | Are your study participants or game players aware that their data will be collected and used? | Y | N |
| 26 | Does your game or app contain addictive design elements without benefit to the user? | N | N |
| 27 | Does your survey contain an aspect of compulsion or unusually large incentive, that would command users to take it even if it was to their detriment? | N | N |
| 28 | Could your research outcomes harm an individual or entity? | N | N |

# Approvals

The signatures of the people below indicate an understanding of the purpose and content of this document by those signing it. By signing this document, you indicate that you approve of the proposed project outlined in this Statement of Work, the division of work, the Ground Rules and that the next steps may be taken to create a Product Specification and proceed with the project.

|  |  |  |  |
| --- | --- | --- | --- |
| Approver Name | Title | Signature | Date |
| Singh Gagan Preet | Project Manager | A close-up of a signature  AI-generated content may be incorrect. | 09/26/2025 |
| Pant Divyansh | Team Member |  | 09/26/2025 |
| Tiwari Sourav | Team Member |  | 09/26/2025 |
| Valani Rameshkumar Premji | Team Member |  | 09/26/2025 |
| Dr. Xiao Hu | Advisor |  |  |

# Appendix

## Advisor Engagement

## Project Team Responsibilities

* The Project Manager will set up and facilitate a weekly call/meeting with the Faculty Advisor. The Project Team will provide weekly status updates to the Faculty Advisor including upcoming deliverables, critical issues, and any adjustments to the Project Plan.
* Documents will be provided to the Faculty Advisor with adequate time for review and signature. The time necessary for review will be agreed with the Advisor. The minimum review time will be 3 days prior to the document due date.
* Design files will be provided to the Faculty Advisor as requested in a format agreed to with the Advisor.
* Support requirements will be clearly requested from the Faculty Advisor with the dates required and an adequate time for fulfilling the request.
* Modifications requests to the Project Plan by Faculty Advisor will be reviewed and agreed to within 1 week of the request.

## Faculty Advisor Responsibilities

* The Faculty Advisor will provide knowledge and expertise to help the group stretch their skills.
* The Faculty Advisor will participate in a weekly or bi-weekly call/meeting with the Project Team to review the project status, upcoming deliverables, priorities, issues, and progress to the agreed Project Plan.
* The Faculty Advisor will provide document review, feedback and approval, rejection, approval with contingencies with adequate time for the Project Team to meet the course due dates.
* The Faculty Advisor will provide feedback to requested support requirements from the Project Team. This includes feedback and guidance on design implementations decisions, design files, test plans, test procedures and test results.
* The Faculty Advisor shall provide technical advice and guidance to the Project Team answering inquiries approximately 1 hour per week.
* Modifications to the Project Plan by the Project Team will be resolved and documented within 1 week of the request.
* Grade the finalized project using a skill-based rubric
* Attend iShowcase in December.

## Ground Rules

As a team and as individual team members, we agree to:

1. **Stay focused on our objectives and goals.**

Each time the team meets, we will clearly define our objectives and desired outcomes at the beginning of the meeting. We will politely remind team members if we are getting off track.

1. **“Sidebar” any issues that are relevant but not consistent with the immediate objectives.**

Occasionally, important matters are raised that are not relevant to the immediate goals of the meeting. To keep the group on track, but avoid losing the issue, create a “sidebar” where these topics can be listed and discussed later.

1. **Listen when others are speaking.**

We will listen and consider others’ input before adding our own comments.

1. **All viewpoints will have an opportunity to be heard.**

We understand that some team members may be quieter than others. We will make an effort to get each team member’s viewpoint and that no one dominates the discussion.

1. **Differences of opinion will be discussed respectfully**

We will identify areas of agreement before assessing areas of disagreement. We will encourage each other to look beyond our own point of view. We will discuss different ideas respectfully. As a team, we will weigh the merits of different opinions and agree on a process for choosing a direction. All team members will respect and follow the decision or direction.

1. **Look for the good points in new ideas.**

We will endeavor to explore the value in each idea as we assess and select our path forward.

1. **Focus on the future, not the past.**

We will use our past experience to inform our decisions, but focus the discussion on the future

objectives. Blame for past performance is counterproductive, we will focus on finding solutions.

1. **Agree upon specific action items and next steps.**

At the end of each meeting and discussion, we will summarize and agree on specific next steps, action items and assignments.

1. **Accountability**

As team members, we will each be responsible for our individual assignments and contribution to achieving the team objectives and goals. We will honor our responsibilities and not let our team members down.