

F25-314-D-CoWriteIA

Project Team

Aisha Siddiqa 22I-1281
Ayesha Ejaz 22I-0899
Junaid Asrar 22I-0770

Session 2022-2026

Supervised by

Dr. Ali Zeeshan



Department of Computer Science

**National University of Computer and Emerging Sciences
Islamabad, Pakistan**

June, 2026

Contents

1	Introduction	1
1.1	Existing Solutions	1
1.2	Problem Statement	2
1.3	Scope	2
1.4	Modules	3
1.4.1	Module 1: Project Indexing and Semantic Memory	3
1.4.2	Module 2: Context-Aware Writing Assistant	3
1.4.3	Module 3: Character and Scene Management	3
1.4.4	Module 4: Dialogue Generation	3
1.4.5	Module 5: Research Integration	3
1.4.6	Module 6: Project Query Interface	3
1.4.7	Module 7: Gap Analysis Module	3
1.5	Work Division	5
2	Project Requirements	7
2.1	Use-case	7
2.2	Functional Requirements	8
2.2.1	Module 1: Project Indexing and Semantic Memory	8
2.2.2	Module 2: Context-Aware Writing Assistant	8
2.2.3	Module 3: Character and Scene Management	8
2.2.4	Module 4: Dialogue Generation	9
2.2.5	Module 5: Research Integration Module	9
2.2.6	Module 6: Project Query Interface	9
2.2.7	Module 7: Gap Analysis Module	9
2.3	Non-Functional Requirements	9
2.3.1	Usability	9
2.3.2	Performance	10
2.3.3	Security	10
2.3.4	Reliability	10
2.3.5	Maintainability	10
2.3.6	Compatibility	10
2.3.7	Scalability	10

3 System Overview	11
3.1 Architectural Design	11
3.2 Data Design	13
3.2.1 Class Diagram	14
3.3 Domain Model	14
3.3.1 Relationships and Associations	16
3.3.2 Domain Model Diagram	17
3.4 Design Models	17
3.4.1 State Transition Diagram	17
 4 Implementation and Testing	19
4.1 Algorithm Design	19
4.1.1 High-Level Algorithm Flow	19
4.1.2 Document Chunking Algorithm	20
4.1.3 Embedding Generation and Storage	20
4.1.4 Context Retrieval Algorithm (RAG)	20
4.1.5 AI Response Generation Algorithm	21
4.1.6 Performance Considerations	21
4.1.7 Reliability and Fault Handling	21
4.2 External APIs and SDKs	22
4.3 Testing Details	22
4.3.1 Black Box Testing	22
4.3.1.1 Purpose of Black Box Testing	22
4.3.1.2 Testing Environment	22
4.3.1.3 Functional API Coverage	22
4.3.1.4 Workflow and Scenario Testing	23
4.3.1.5 Negative and Security Testing	23
4.3.1.6 Error and Edge Case Validation	23
4.3.1.7 Black Box Test Evidence	23
4.3.1.8 Black Box Testing Summary	24
4.3.2 Unit Testing	24
4.3.2.1 Implemented Unit Test Coverage	24
4.3.2.2 Testing Approach	24
4.3.2.3 Failure and Boundary Validation	25
4.3.3 Test Evidence	25
4.4 Test Summary	25
 5 Conclusions and Future Work	27
5.1 Conclusion	27
5.2 Future Work	27

References	29
A Appendices	30
A.1 Appendix A: System Diagrams	30
A.1.1 Use Case Diagram	30
A.1.2 Activity Diagram	31
A.2 Appendix B: Architecture Diagram	32
A.2.1 CoWriteIA Detailed Architecture Diagram	32
A.3 Appendix C: User Interface Screenshots	33
A.3.1 Dashboard Interface	33
A.3.2 Writing Environment Interface	33
A.3.3 Writing Assistant Interface	34
A.3.4 Login Interface	34
A.4 Appendix D: Detailed Use Case Specification	36
A.4.1 UC-01: Generate Context-Aware Writing	36

List of Figures

1.1	Gantt Chart - Project Timeline	6
3.1	Box and Line Diagram	13
3.2	Entity–Relationship Diagram for CoWriteIA	14
3.3	Class Diagram for CoWriteIA	15
3.4	Domain Model Diagram for CoWriteIA	17
3.5	State Transition Diagram for CoWriteIA	18
4.1	BlackBox Tests fig:1	23
4.2	BlackBox Tests fig:2	24
4.3	Test Execution Results	25
A.1	Use Case Diagram CoWriteIA	30
A.2	Activity Diagram	31
A.3	CoWriteIA Multi-Tier Architecture Diagram	32
A.4	CoWriteIA Dashboard Interface	33
A.5	Writing Environment Interface	33
A.6	Writing Assistant Interface	34
A.7	Login Interface	34

List of Tables

1.1	Comparison of Existing Solutions	2
1.2	Work Division for Iteration I	5
1.3	Work Division for Iteration II	5
3.1	Architectural Tiers and Responsibilities	12
4.1	Document Chunking Algorithm	20
4.2	Embedding Generation Algorithm	20
4.3	Context Retrieval Algorithm (RAG)	21
4.4	AI Response Generation Algorithm	21
4.5	External APIs and SDKs Used	22
A.1	Detailed Use Case: Generate Context-Aware Writing	36

Chapter 1

Introduction

CoWriteIA is an AI-powered writing assistant designed to support creative writers, novelists, researchers, and content creators. Modern writing workflows require managing notes, drafts, characters, scenes, and references across several disconnected tools, which often breaks focus and reduces productivity. Writers struggle to maintain narrative consistency, track earlier ideas, and keep a coherent writing style when working on long projects. Existing tools offer only isolated features such as grammar correction or basic generation and do not provide project-level understanding or semantic memory [5, 8].

CoWriteIA addresses these issues by creating a unified, intelligent workspace where all project files are indexed, searchable, and semantically connected. By integrating project-level memory, context-aware writing, dialogue generation, character management, and research support, the system helps writers maintain consistency and improve their creative process. The platform is designed to reduce cognitive load, avoid fragmented workflows, and provide meaningful AI assistance throughout the writing journey. This chapter presents the background, existing solutions, problem statement, scope, project modules, and work division that form the foundation of this system.

1.1 Existing Solutions

Several writing and AI-assisted tools exist, but each focuses on limited aspects of the writing process. Grammarly offers grammar correction and style suggestions but lacks deep contextual awareness across long projects. Notion AI and Jasper AI provide generative assistance but do not maintain story-level continuity or user-specific writing style. Tools like Scrivener help with organization but have no semantic understanding or AI memory. As a result, writers must repeatedly switch between applications to manage notes, drafts, characters, and research, leading to inefficiency and inconsistent writing flow.

Table 1.1: Comparison of Existing Solutions

System Name	System Overview	System Limitations
Grammarly	Provides grammar correction, clarity improvements, and tone suggestions.	No project awareness, no memory of earlier chapters, no semantic search.
Notion AI	Offers AI-assisted content generation and note organization.	Cannot maintain narrative consistency; lacks dialogue and character support.
Scrivener	Strong organizational tool for large writing projects with chapter/scene structure.	No AI support, no semantic retrieval, no automated gap or style analysis.

1.2 Problem Statement

Writers working on long-form projects often lose track of earlier ideas, character traits, plot points, and stylistic decisions. This leads to inconsistencies, repeated ideas, and a break in narrative flow. Existing tools either provide isolated writing support or document organization but do not combine both with meaningful context. AI tools can generate text but fail to maintain project-level continuity, making the generated text feel disconnected from the writer’s established style or storyline. Writers spend significant time searching through older drafts to recall information [3, 5].

CoWriteIA aims to solve these problems by offering an intelligent system that continuously indexes all project content, retrieves relevant context, and assists writers in generating text that aligns with their established narrative and writing style. By maintaining a unified knowledge base and supporting character consistency, scene management, dialogue generation, and semantic search, the system reduces cognitive overhead and improves creative flow.

1.3 Scope

The scope of CoWriteIA includes building an AI-driven writing environment that supports project management, semantic search, context-aware writing, dialogue generation, research integration, and style adaptation. The system will allow users to upload documents, create new content, store character information, generate dialogues, and retrieve context from a vector-based semantic memory [4, 8]. It will support long-form writing projects such as novels, research documents, and story-driven content.

The system will not include plagiarism detection, multimedia editing, or full publishing workflows. It focuses strictly on improving the writing process, maintaining consistency, and providing intelligent assistance throughout the creative workflow.

1.4 Modules

The project consists of several modules, each responsible for a unique part of the writing workflow.

1.4.1 Module 1: Project Indexing and Semantic Memory

This module extracts, embeds, and organizes all project content into a semantic database for context-aware retrieval [8].

1. Automatic project indexing and embedding generation.
2. Semantic search based on meaning instead of keywords.

1.4.2 Module 2: Context-Aware Writing Assistant

This module provides intelligent writing suggestions that match the user's tone and project context [5].

1. Generates coherent drafts aligned with past content.
2. Retrieves relevant information to maintain consistency.

1.4.3 Module 3: Character and Scene Management

This module stores and manages characters, scenes, and narrative details.

1. Character profiles with traits and relationships.
2. Scene storage and tracking.

1.4.4 Module 4: Dialogue Generation

Generates natural, character-consistent dialogue suggestions.

1. Dialogue generation based on personality and context.
2. Supports narrative flow within scenes.

1.4.5 Module 5: Research Integration

Fetches factual data from external sources for realistic writing.

1. Web-based research retrieval.
2. Insertable factual references.

1.4.6 Module 6: Project Query Interface

Allows users to ask natural-language questions about their own project.

1. Semantic question-answering.
2. Source-linked responses.

1.4.7 Module 7: Gap Analysis Module

Analyzes draft content to find missing or weak areas.

1. Introduction

1. Identifies incomplete sections.
2. Suggests improvements for clarity and consistency.

1.5 Work Division

The work completed during FYP-1 was divided into two major iterations. A summary of responsibilities is presented in the tables below.

Iteration I Tasks

Table 1.2: Work Division for Iteration I

Task	Ayesha	Junaid	Aisha
SRS Document	✓	✓	✓
UML Diagrams	✓	✓	✓
UI Design		✓	✓
Database Connection		✓	✓
Frontend (Main Pages)	✓		
Project Indexing Agent	✓	✓	
Knowledge Retrieval Agent		✓	✓
ChatBot Support		✓	

Iteration II Tasks

Table 1.3: Work Division for Iteration II

Task	Ayesha	Junaid	Aisha
Inline Copilot Support	✓		
Context-Aware Writing Module	✓		
Gap Analysis Module			✓
Style Adaptation		✓	
Testing	✓	✓	✓
Documentation (All Sections)	✓	✓	✓

Project Timeline

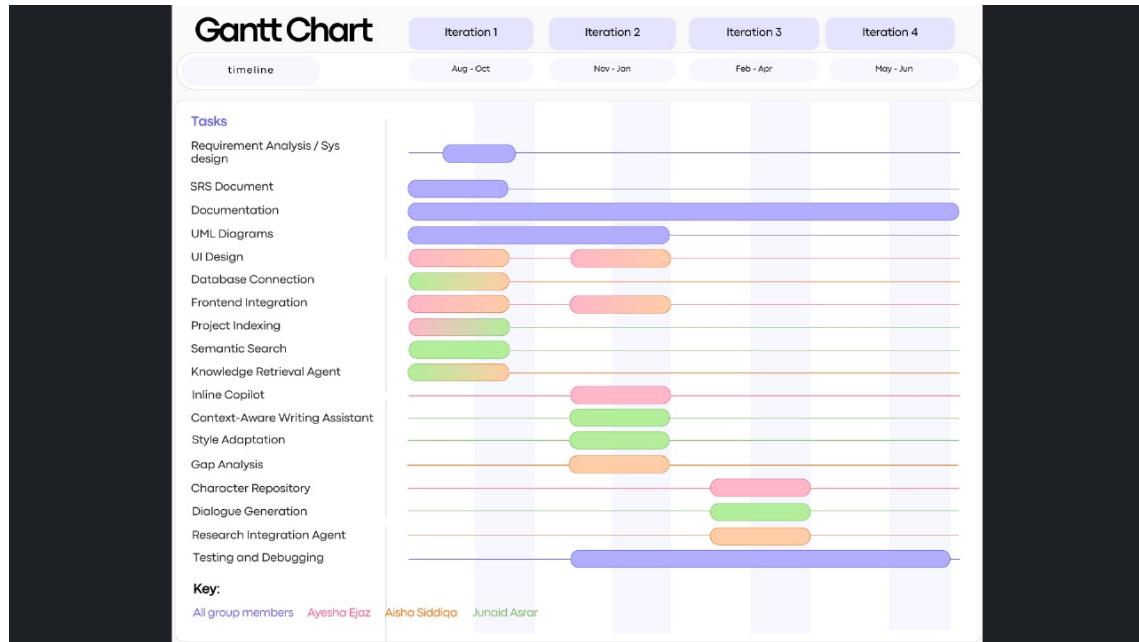


Figure 1.1: Gantt Chart - Project Timeline

Chapter 2

Project Requirements

This chapter defines the requirements essential for developing CoWriteIA. These requirements were derived from the FYP-1 Proposal, Mid Report, and the SRS document. They describe how the system should behave, how users will interact with it, and what constraints and quality standards must be followed. The requirements are divided into functional and non-functional categories, with additional details about expected user interactions.

2.1 Use-case

CoWriteIA provides writers with AI-assisted features such as project indexing, semantic retrieval, character management, context-aware writing, and dialogue generation. To clearly define how users interact with these features, system behavior is modeled through detailed use cases.

A use case describes the interaction between the primary actor (the writer) and the system to accomplish a specific task. Each use case outlines the trigger event, preconditions, main workflow steps, and system responses. These models help ensure that system functionality aligns with user expectations and guide the design of interactive features.

In addition to textual use cases, a visual representation in the form of a Use-Case Diagram provides an overview of the major user interactions with CoWriteIA. This includes actions such as uploading documents, retrieving semantic context, generating content with the writing assistant, managing characters, and initiating dialogue generation. Such diagrams help maintain clarity, consistency, and completeness throughout the system design.

A detailed use case example for the core functionality **Generate Context-Aware Writing (UC-01)** is provided in Appendix D. This detailed use case outlines the complete interaction flow including preconditions, main flow steps, alternate flows, exceptions, and postconditions. The use case demonstrates how the system retrieves semantic context, processes user requests through the AI model, and delivers generated content while han-

dling various edge cases such as missing embeddings or service unavailability.

2.2 Functional Requirements

The functional requirements describe the operations the CoWriteIA system must support. These requirements come directly from the system features identified in the Proposal and SRS.

2.2.1 Module 1: Project Indexing and Semantic Memory

This module handles ingestion, indexing, and semantic storage of project documents.

1. The system shall allow users to upload project files including text documents, chapters, notes, and research material.
2. The system shall extract, segment, and convert uploaded content into embeddings for semantic retrieval.
3. The system shall store embeddings in a vector database.
4. The system shall provide semantic search capabilities based on meaning rather than keyword matching.
5. The system shall return ranked search results with source references.

2.2.2 Module 2: Context-Aware Writing Assistant

This module generates content aligned with user writing style and project context.

1. The system shall analyze previous project content to understand tone, terminology, and style.
2. The system shall allow users to generate context-aware text suggestions based on previous chapters or notes.
3. The system shall retrieve relevant context automatically when generating new content.
4. The system shall maintain style consistency between newly generated and existing content.

2.2.3 Module 3: Character and Scene Management

This module manages characters, their traits, and related narrative structures.

1. The system shall allow users to create, edit, and store character profiles.
2. The system shall store attributes such as personality, relationships, behaviors, and backstory.
3. The system shall allow users to manage scenes and attach characters to scenes.
4. The system shall assist in retrieving character information when generating story text or dialogue.

2.2.4 Module 4: Dialogue Generation

This module generates consistent, character-matching dialogue.

1. The system shall generate dialogue aligned with character personality and tone.
2. The system shall allow users to request dialogue for specific characters or scenes.
3. The system shall ensure continuity between generated dialogue and existing narrative.

2.2.5 Module 5: Research Integration Module

This module retrieves factual information to support realistic writing.

1. The system shall allow users to search for factual references.
2. The system shall fetch external information from trusted research sources.
3. The system shall present research results with citations.

2.2.6 Module 6: Project Query Interface

This module allows users to ask natural-language questions about their project.

1. The system shall process user queries related to characters, scenes, chapters, or events.
2. The system shall fetch relevant information from the semantic memory.
3. The system shall provide answers with linked source passages.

2.2.7 Module 7: Gap Analysis Module

This module identifies missing or weak areas of the writer's draft.

1. The system shall analyze uploaded chapters for missing elements such as incomplete scenes or inconsistent character behavior.
2. The system shall highlight areas needing expansion or clarification.
3. The system shall provide suggestions to improve narrative flow and completeness.

2.3 Non-Functional Requirements

Non-functional requirements ensure that CoWriteIA performs reliably, efficiently, and securely.

2.3.1 Usability

1. The system shall provide a simple and intuitive interface accessible to writers with minimal technical expertise.
2. The system shall allow users to perform core actions such as uploading files, searching, and generating text within no more than three interactions.
3. The UI shall clearly present semantic search results with source references.

2.3.2 Performance

1. The system shall index uploaded documents within 5 seconds for an average chapter-length file.
2. Semantic search results shall appear within 2 seconds of a query.
3. Generated text responses shall be produced within 3–5 seconds depending on context length.

2.3.3 Security

1. User project data shall be stored securely using encrypted connections.
2. Only authenticated users shall access their own documents.
3. The system shall not use project data for training without user permission.

2.3.4 Reliability

1. The system shall maintain uptime of at least 99%
2. The system shall recover gracefully from API or model failures by retrying or presenting fallback responses.

2.3.5 Maintainability

1. The codebase shall follow modular architecture to allow updates to individual agents.
2. The system shall support integration of new language models without requiring major structural changes.

2.3.6 Compatibility

1. The system shall run on modern browsers including Chrome, Edge, and Firefox.
2. The frontend shall be built using Next.js and shall be compatible with desktop and tablet interfaces.

2.3.7 Scalability

1. The vector database shall support scaling as the user creates larger projects.
2. The system shall handle multiple concurrent requests without significant performance degradation.

Chapter 3

System Overview

Introduction

This chapter provides a high-level description of the system's overall functionality, context, and architectural design. The aim is to present how the major parts of the system interact and why the system has been decomposed into specific modules and tiers.

System Overview

CoWriteIA is an AI-assisted writing platform designed to support writers through various stages of the creative process [2, 6, 7]. The system enables users to:

- Create and manage writing projects
- Upload and organize documents
- Manage character profiles and story elements
- Generate and refine written content using AI assistance

To efficiently support these features, the system separates concerns into distinct, coordinated components: user interaction, application logic, AI processing, and data management.

3.1 Architectural Design

The architecture of CoWriteIA follows a **multi-tier model**, combining **client-server** and **layered** principles. This structure organizes the system into four logical tiers, each with specific responsibilities to ensure maintainability, scalability, and clarity.

Architectural Tiers

Data Storage Components

- **Main Database:** Stores structured system data (users, projects, characters)
- **Vector Database:** Manages embeddings for semantic retrieval and search

Tier	Component	Responsibilities
Presentation	Frontend Application	User interaction, interface rendering, and user input handling
Application Logic	Backend (API Server)	Authentication, business workflows, and project/document/character management
AI Processing	Processing Worker	Embeddings generation, semantic retrieval, and AI text generation
Data	Storage Systems	Structured data, vector storage, and file management

Table 3.1: Architectural Tiers and Responsibilities

- **File Storage:** Handles uploaded documents and exported files

Architectural Diagram

A detailed architecture diagram illustrating the complete multi-tier structure of CoWriteIA is provided in Appendix C (Figure A.3).

Diagram Development Process

- **Initial Design Stage:** Create a *Box and Line Diagram* for simpler representation of systems
- **Finalization Stage:** After selecting the architecture style/pattern (MVC, Client-Server, Layered, Multi-tiered), create detailed mapping of modules/components to each part of the architecture

Design Principles

This architectural decomposition supports the system's functional requirements while ensuring:

- **Flexibility:** Components can be modified independently
- **Maintainability:** Clear separation of concerns simplifies updates and debugging
- **Scalability:** Each tier can be scaled independently based on demand
- **Security:** Controlled data flow between tiers with proper authentication

Key Architectural Decisions

- **Separation of AI Processing:** Intensive AI tasks are handled by dedicated workers to maintain API responsiveness
- **Multi-tier Structure:** Enables independent development, testing, and deployment of each tier

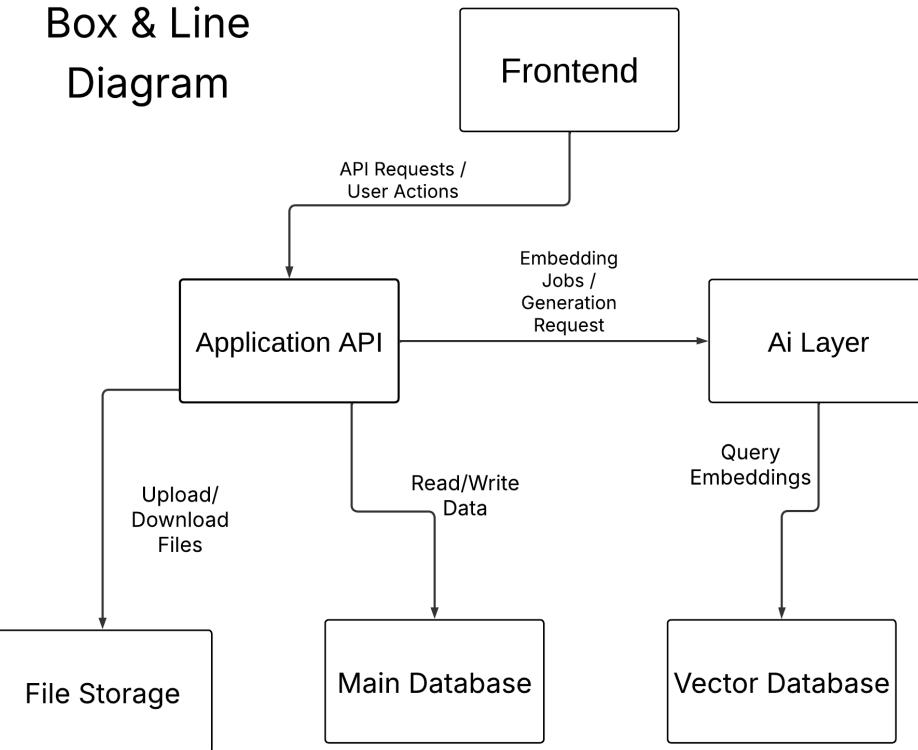


Figure 3.1: Box and Line Diagram

- **Modular Design:** Supports future enhancements and feature additions

3.2 Data Design

The data design of CoWriteIA transforms functional requirements into structured data models that support **project creation**, **AI-assisted writing**, **semantic search**, **character management**, and **dialogue modelling**.

Data Organization

The system organizes data into three primary categories:

- **Relational Data:** User accounts, projects, documents, characters, and chat sessions stored in a structured database with defined schemas and foreign key relationships
- **Vector Data:** High-dimensional embeddings generated from text content, enabling semantic similarity search and context retrieval
- **Binary Data:** Uploaded files and exported documents stored separately with metadata linkage

Data Flow and Transformation

Data flows through multiple transformation stages:

- **Input Processing:** User actions validate and create/update structured database records
- **Embedding Generation:** Text content is transformed into vector representations for AI operations
- **Retrieval Augmentation:** Semantic queries fetch relevant context from vector and relational stores
- **Output Generation:** AI responses and exports combine retrieved data with generated content

This design ensures **data integrity**, **efficient retrieval**, and **scalable processing**. Recent work on efficient local models such as Phi-3 highlights practical on-device capabilities for embedding and retrieval [1].

Figure 3.2 provides the Entity–Relationship Diagram showing database schema structure, while Figure 3.4 presents the conceptual domain model.

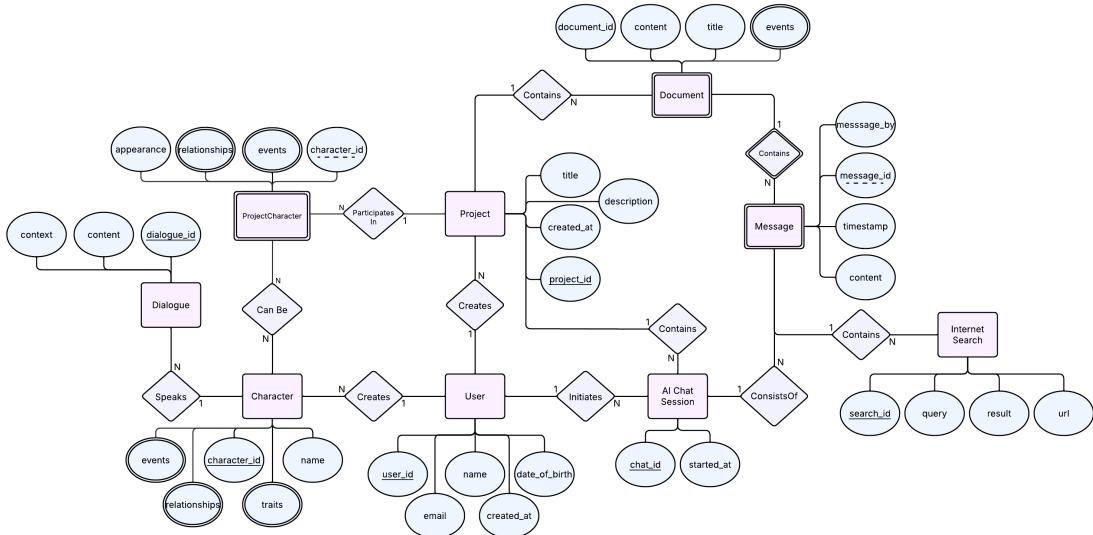


Figure 3.2: Entity–Relationship Diagram for CoWriteIA

3.2.1 Class Diagram

The class diagram represents the object-oriented view of the system and shows classes, attributes, and associations used by the application logic. It provides a structural foundation for backend implementation and helps ensure consistency between the conceptual design and the code-level architecture.

3.3 Domain Model

The domain model provides a conceptual view of the system's core entities and their relationships, bridging functional requirements and technical implementation. It describes

Class Diagram

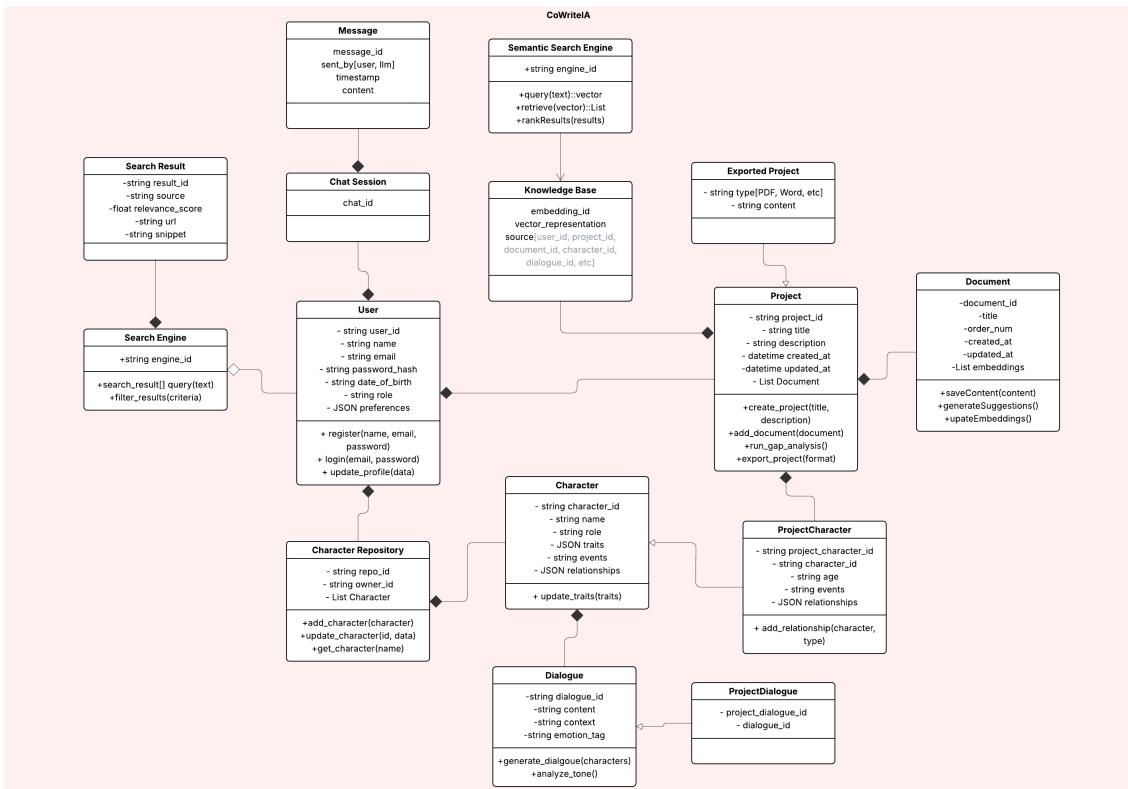


Figure 3.3: Class Diagram for CoWriteIA

the key business objects, their attributes, and how they interact within CoWriteIA.

Entities and Attributes

The major entities of the system include:

- **User:** Represents a system user who owns projects and interacts with the AI. Attributes include `user_id`, `name`, `email`, `date_of_birth`, `password_hash`, `role`, and `created_at`.
- **Project:** A workspace created by the user. Attributes include `project_id`, `title`, `created_at`, and `updated_at`.
- **Document:** Represents a chapter or section of a project. Attributes include `document_id`, `title`, `order_num`, `created_at`, and `updated_at`.
- **Character Repository:** A user-owned collection that groups all character profiles.
- **Character:** Represents a story character. Attributes include `character_id`, `name`, `traits`, `history`, `created_at`, and `updated_at`.
- **ProjectCharacter:** An association class linking characters with specific projects. Stores project-specific values such as `age` and `events`.
- **Dialogue:** Represents character dialogues. Attributes include `dialogue_id`, `content`, `context`, and `emotion_tag`.
- **ProjectDialogue:** A linking entity that associates dialogues with a project.
- **Chat Session:** Represents an AI chat interaction initiated by the user. Identified by `chat_id`.
- **Message:** Stores an individual message in a chat session. Attributes include `message_id`, `sent_by`, `timestamp`, and `content`.
- **Search Engine:** Represents the semantic search component.
- **Search Result:** Stores contextual search outputs. Attributes include `result_id`, `query`, and `content`.
- **Knowledge Base:** Stores vector embeddings and source references. Attributes include `embedding_id`, `vector_representation`, and `source_ids`.
- **Exported Project:** Represents generated output files such as PDF or Word exports.

3.3.1 Relationships and Associations

The domain model defines the relationships that structure how information flows across the system:

- A **User** owns multiple **Projects**.
- A **Project** contains multiple **Documents**.
- A **User** owns a **Character Repository**, which stores many **Characters**.
- A **Project** includes multiple **Characters** via **ProjectCharacter**.
- A **Character** speaks one or more **Dialogues**.
- A **Project** associates specific dialogues through **ProjectDialogue**.
- A **User** initiates a **Chat Session**.

- A **Chat Session** stores multiple **Messages**.
- The **LLM Chatbot** interacts with the **Semantic Search Engine** and **Knowledge Base** to generate responses.
- The **Search Engine** produces multiple **Search Results**, which may be injected into messages or documents.

These relationships ensure that writing, character creation, semantic search, and AI interactions remain consistent and interconnected across the system.

3.3.2 Domain Model Diagram

Figure 3.4 presents the complete domain model for CoWriteIA. It includes all major entities, attributes, and associations, including weak entities and conceptual relationships.

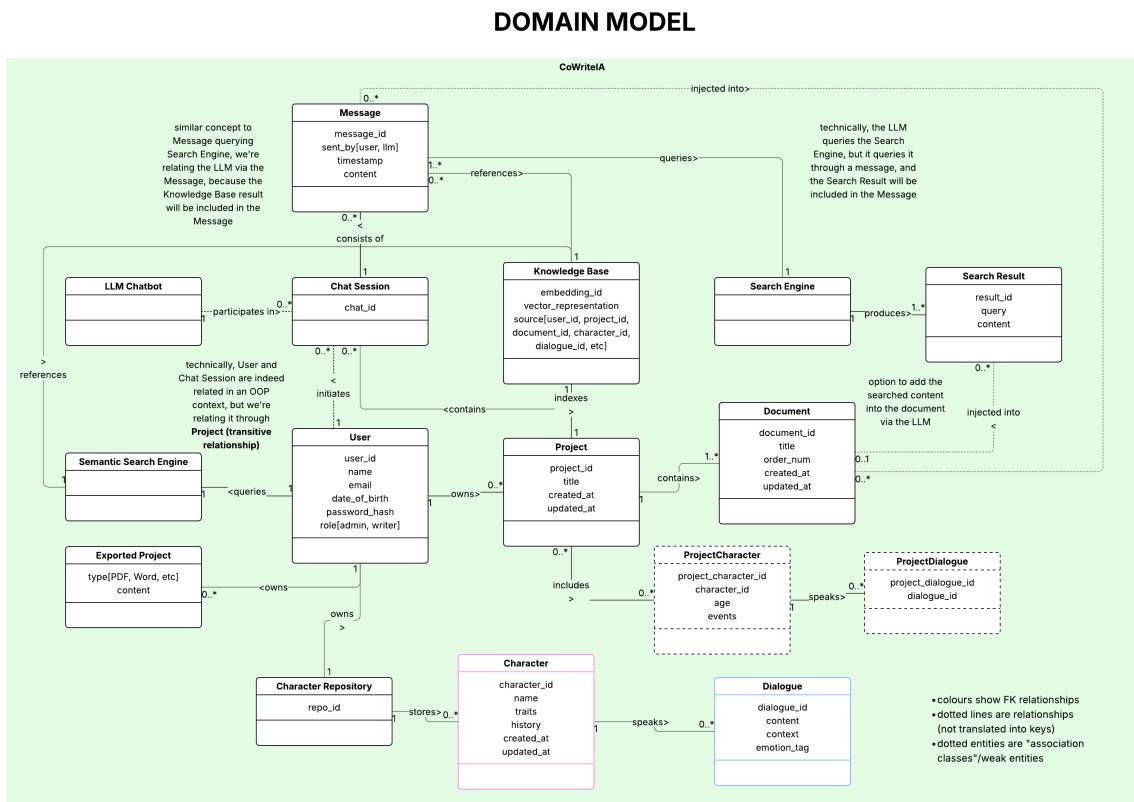


Figure 3.4: Domain Model Diagram for CoWriteIA

3.4 Design Models

3.4.1 State Transition Diagram

Figure 3.5 illustrates state transitions for key entities based on user actions and system events. The diagram shows how entities move between states:

- **Project**: Transitions from Draft → Active → Archived/Deleted based on user ac-

tions

- **Document:** Moves through Creating → Editing → Saved → Publishing → Published states
 - **Chat Session:** Cycles between Idle → Processing → Waiting for User → Completed
 - **Character:** Changes from Draft → Active → Archived as needed

These transitions ensure proper lifecycle management and data consistency throughout the system.

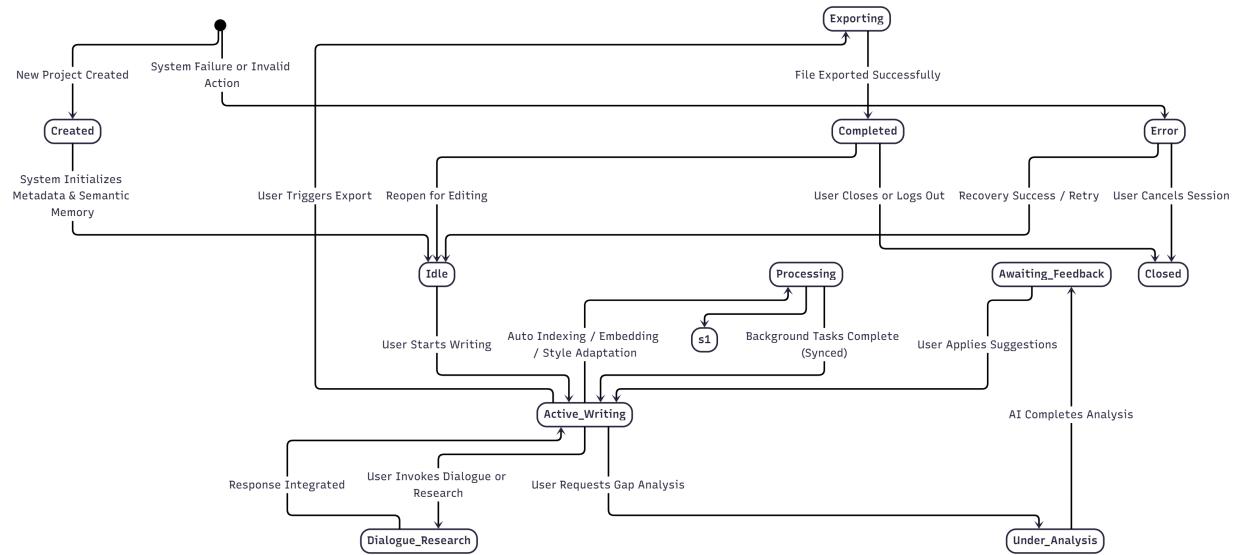


Figure 3.5: State Transition Diagram for CoWriteIA

Chapter 4

Implementation and Testing

This chapter describes the current implementation status of the CoWriteIA system and the testing strategies applied to validate its functionality. The focus of this phase was to implement the core backend services, major APIs, and conduct rigorous black box testing to ensure that all user-facing functionalities behave correctly according to the system requirements.

The implementation described in this chapter reflects the progress achieved up to the current iteration of the project.

4.1 Algorithm Design

The system implements a Retrieval-Augmented Generation (RAG) pipeline to enable intelligent document-based responses. The algorithm focuses only on the core functional requirements: document processing, semantic indexing, context retrieval, and AI response generation.

4.1.1 High-Level Algorithm Flow

The overall system follows this sequence:

1. Document upload by the user
2. Text extraction and chunking
3. Embedding generation for semantic indexing
4. Vector storage
5. Query-based context retrieval
6. AI-based response generation

This structured flow enables efficient and relevant AI-powered responses.

4.1.2 Document Chunking Algorithm

The system divides uploaded documents into smaller logical chunks. This improves semantic accuracy and ensures efficient storage.

Table 4.1: Document Chunking Algorithm

Pseudocode

Algorithm: Document Chunking
Input: Full document text
Output: List of text chunks
Split the document into sentences.
Initialize an empty list of chunks.
Append sentences to a chunk until size limit is reached.
Store the completed chunk and start a new one.
Store remaining text as the final chunk.

4.1.3 Embedding Generation and Storage

Each text chunk is converted into a numerical vector (embedding) that represents semantic meaning. These vectors are stored in a vector database to allow fast similarity search.

Table 4.2: Embedding Generation Algorithm

Pseudocode

Algorithm: Embedding Generation
Input: List of text chunks
Output: Vector embeddings
Read each text chunk.
Convert text chunks into semantic vector embeddings.
Store embeddings in vector database.
Link embeddings with source chunks.

4.1.4 Context Retrieval Algorithm (RAG)

When a user submits a query, the system retrieves the most relevant chunks using semantic similarity search.

Table 4.3: Context Retrieval Algorithm (RAG)

Pseudocode

Algorithm: Context Retrieval (RAG)
Input: User query
Output: Relevant text chunks
Generate embedding for the user query.
Search vector database for similar embeddings.
Select top-k most relevant chunks.
Return these chunks as context.

4.1.5 AI Response Generation Algorithm

The system combines the retrieved context with the user query and sends it to the language model to generate accurate, context-aware responses.

Table 4.4: AI Response Generation Algorithm

Pseudocode

Algorithm: AI Response Generation
Input: User query and retrieved context
Output: AI generated response
Combine query and contextual information.
Format the data into a prompt.
Send the prompt to the language model.
Receive and return the generated response.

4.1.6 Performance Considerations

- Chunking runs in linear time relative to document size.
- Embedding generation dominates processing time.
- Vector search operates in sub-linear time using approximate nearest neighbor indexing.

4.1.7 Reliability and Fault Handling

The system includes fallback mechanisms to handle:

- Missing embeddings
- Vector database unavailability
- Invalid user input

These ensure stable system behavior under failure scenarios.

4.2 External APIs and SDKs

The system integrates multiple third-party APIs and SDKs to support AI capabilities, authentication, and data persistence.

API / SDK	Description	Purpose of Usage	Endpoint/ Function
OpenAI API (v1)	Large language model services	Text generation and embeddings	/v1/chat/completions
FastAPI	Python web framework	REST API backend	@app.get(), @app.post()
PostgreSQL	Relational database system	User and project data storage	psycopg2 driver
VectorDB (FAISS/Pinecone)	Vector similarity engine	Semantic search and retrieval	similarity_search()

Table 4.5: External APIs and SDKs Used

4.3 Testing Details

Testing played a critical role in validating the correctness, reliability, and stability of the implemented system. A combination of black box testing and unit testing strategies was adopted to ensure that the system meets its functional requirements.

4.3.1 Black Box Testing

Black box testing was used to validate the external behavior of the system without reference to internal implementation details. The focus was on verifying that system features produced correct outputs when interacting through public interfaces such as API endpoints.

4.3.1.1 Purpose of Black Box Testing

The objective of this testing was to ensure that system functionality aligned with the documented functional requirements by validating response codes, output structures, and observable system behavior.

4.3.1.2 Testing Environment

Testing was conducted using an isolated staging environment that included a running API server, a connected database, and a REST API testing client. All tests were executed externally without accessing source code.

4.3.1.3 Functional API Coverage

The following functional modules were tested as black box components:

- Authentication services (registration, login, token validation)
- Project management operations
- File handling operations
- Search and retrieval services
- Chat and conversational APIs

4.3.1.4 Workflow and Scenario Testing

End-to-end user workflows were validated through multi-step test scenarios, including project creation, file indexing, semantic search, and AI-assisted responses.

4.3.1.5 Negative and Security Testing

Invalid inputs, unauthorized requests, and forbidden access attempts were tested to verify that the system safely rejected improper usage without exposing internal errors.

4.3.1.6 Error and Edge Case Validation

Boundary conditions such as missing data, empty result sets, and invalid resource requests were evaluated to ensure stable system responses.

4.3.1.7 Black Box Test Evidence

All executed test cases were recorded and maintained as structured documentation.

	A	B	C	D	E	F	G	H	I	J
1	TC ID	Module	Test Case Name	Description	Test Type	Preconditions	Steps / Input	Expected Result	Priority	Tags / Req ID
2	TC-01	System Health	Root Endpoint	Verify root endpoint	Black Box	API server running	GET /	200 OK, valid response	High	health
3	TC-02	System Health	Health Check API	Verify /health endpoint	Black Box	API server running	GET /health	200 OK, status = "OK"	High	health
4	TC-03	System Health	API Docs	Verify API docs available	Black Box	API server running	GET /docs or /openapi	200 OK, docs available	Medium	docs
5	TC-04	Authentication	User Registration	Register user with valid credentials	Black Box	DB online, user not exists	POST /auth/register	201 Created, user added	High	REQ-auth-1
6	TC-05	Authentication	User Registration	Register user with existing email	Black Box	User already exists	POST /auth/register	400/409 error response	High	negative
7	TC-06	Authentication	Login (Valid)	Login with correct credentials	Black Box	Registered user	POST /auth/login	200 OK, access token issued	High	token
8	TC-07	Authentication	Login (Invalid)	Login with wrong credentials	Black Box	User exists	POST /auth/login	401 Unauthorized	High	negative
9	TC-08	Authentication	Get Current User	Retrieve user profile	Black Box	Valid access token	GET /auth/me	200 OK, user data returned	Medium	auth
10	TC-09	Authentication	Logout	Invalidate user session	Black Box	Logged in session	POST /auth/logout	200 OK, token invalidated	Medium	auth
11	TC-10	Authentication	Refresh Token	Refresh expired token	Black Box	Valid refresh token	POST /auth/refresh	200 OK, new access token issued	Medium	auth
12	TC-11	Projects	Create Project	Create new project	Black Box	Authenticated user	POST /projects	201 Created, project created	High	REQ-proj-1
13	TC-12	Projects	List Projects	Retrieve projects	Black Box	Multiple projects	GET /projects?per_page=10	200 OK, paginated results	Medium	pagination
14	TC-13	Projects	Get Project Details	Retrieve project details	Black Box	Project exists	GET /projects/{id}	200 OK, correct details	Medium	projects
15	TC-14	Projects	Update Project	Update project details	Black Box	Project exists, authorized	PUT/PATCH /projects/{id}	200 OK, updated successfully	High	projects
16	TC-15	Projects	Delete Project	Delete project	Black Box	Project exists	DELETE /projects/{id}	200/204, project deleted	High	cleanup
17	TC-16	Projects	Project Access Control	Cross-user project access	Black Box	Multiple users/projects	Access other user's project	403 Forbidden	High	security
18	TC-17	Projects	Duplicate Project	Create project with same name	Black Box	Project with same name	POST /projects/{name}	400/409 validation error	Medium	validation
19	TC-18	Files	Upload File	Upload file to project	Black Box	Project exists, authorized	POST /projects/{id}/files	201 Created, file uploaded	High	file-upload

Figure 4.1: BlackBox Tests fig:1

4. Implementation and Testing

	A	B	C	D	E	F	G	H	I	J
30	TC-29	Search	Generate Embedding	Generate vector	Black Box	Model available	POST /embeddir	200 OK, vector returned	Medium	embeddings
31	TC-30	Search	Calculate Similarity	Calculate similarity	Black Box	Vectors exist	POST /similarity	200 OK, similarity	Low	utility
32	TC-31	Search	Find Similar Concepts	Retrieve similar concepts	Black Box	Indexed corpus	POST /similar	200 OK, similar concepts	Medium	search
33	TC-32	Search	Search with Filters	Filtered search	Black Box	Filterable data	POST /search (filter)	200 OK, filtered results	Medium	filters
34	TC-33	Search	Search Paginated	Paginated search	Black Box	Large dataset	POST /search (page)	Paginated responses	Low	pagination
35	TC-34	Search	Search Without Project	Missing project version	Black Box	Invalid/missing project	POST /search	400/422 validation error	High	negative
36	TC-35	Search	Embedding Statistics	Retrieve embeddings	Black Box	Embeddings stored	GET /embedding	200 OK, stats returned	Low	monitoring
37	TC-36	Search	Autocomplete	Suggest search terms	Black Box	Index built	GET/POST /auto	200 OK, suggestion	Low	UX
38	TC-37	Security	Unauthorized Access	Protected API without token	Black Box	No/invalid token	Call secured endpoint	401/403 Unauthorized	Critical	security
39	TC-38	Chat	Send Chat Message	Send message to user	Black Box	Chat service available	POST /chat	200 OK, AI response	Medium	chat
40	TC-39	Chat	Chat With Context	Contextual conversation	Black Box	Chat history exists	POST /chat (with context)	200 OK, context included	Medium	context
41	TC-40	Chat	Get Chat History	Retrieve conversation history	Black Box	Conversation exists	GET /chat/{id}/history	200 OK, history included	Low	history
42	TC-41	E2E	Complete Project	End-to-end project	Black Box	All services running	Full flow execution	Workflow successful	High	e2e
43	TC-42	E2E	File Update Workflow	File update and reindex	Black Box	Existing file	Update file → reindex	Updated content	High	e2e
44	TC-43	E2E	Multi-file Search	Cross-file search	Black Box	Multiple files indexed	Run cross-file search	Correct cross-file	Medium	similarity
45	TC-44	Integration	Comprehensive Integration	Full indexing pipeline	Black Box	NLP + vector DB	Run complete pipeline	Pipeline completed	Low	comprehensive
46	TC-45	Integration	Embedding Integration	Embedding + vector storage	Black Box	Embedding service	Generate and store	Vectors stored & indexed	Medium	integration
47	TC-46	Integration	Async Indexing	Background asynchronous	Black Box	Worker queue active	Trigger async tasks	Tasks completed	Medium	async
48	TC-47	Integration	Relationship Discovery	Entity relationship mapping	Black Box	Extracted entities	Run relationship discovery	Correct relationships	Low	NLP

Figure 4.2: BlackBox Tests fig:2

4.3.1.8 Black Box Testing Summary

The system demonstrated consistent and correct behavior for valid use cases and safely handled invalid or unauthorized operations.

4.3.2 Unit Testing

Unit testing focused on verifying the internal logic of isolated system components. Each unit was tested independently to ensure that business logic, validation rules, and service-level operations functioned correctly.

4.3.2.1 Implemented Unit Test Coverage

Unit tests were implemented for the following components:

- Authentication API (register, login, token handling)
- Projects API (CRUD operations)
- Files API (upload, download, deletion)
- Search API (semantic, hybrid, and filter-based queries)
- Chat API (conversation handling and context management)
- Complete workflows (multi-step user journeys)
- Error cases (400, 401, 403, 404, 500)
- Edge cases (pagination and unauthorized access)

4.3.2.2 Testing Approach

Service functions, repositories, and helper methods were tested in isolation using mocks to remove external dependencies such as databases and third-party services.

4.3.2.3 Failure and Boundary Validation

Tests confirmed that invalid inputs, unauthorized access, and unsupported operations were handled safely without system failures.

4.3.3 Test Evidence

Summary

121 tests took 00:04:32.

(Un)check the boxes to filter the results.

6 Failed, 115 Passed, 0 Skipped, 0 Expected failures, 0 Unexpected passes, 0 Errors, 0 Reruns

Result ▲	Test
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_register_new_user
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_register_duplicate_email
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_register_invalid_password
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_register_password_mismatch
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_login_valid_credentials
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_login_invalid_credentials
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_login_nonexistent_user
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_get_current_user
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_get_current_user_unauthorized
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_get_current_user_invalid_token
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_logout
Passed	tests/test_auth_api.py::TestAuthenticationAPI::test_refresh_token
Passed	tests/test_chat_api.py::TestChatAPI::test_send_chat_message
Passed	tests/test_chat_api.py::TestChatAPI::test_chat_with_context
Passed	tests/test_chat_api.py::TestChatAPI::test_chat_conversation_history
Passed	tests/test_chat_api.py::TestChatAPI::test_chat_unauthorized
Passed	tests/test_chat_api.py::TestChatAPI::test_chat_without_permission

Figure 4.3: Test Execution Results

4.4 Test Summary

The testing results showed that the system successfully handled valid requests and appropriately rejected invalid or unauthorized access attempts. The core workflows such as project creation, file uploading, semantic search, and AI-assisted responses were verified to function as expected during the current iteration.

All critical defects identified during testing were resolved, and remaining enhancements will be addressed in the next development phase.

Chapter 5

Conclusions and Future Work

5.1 Conclusion

The first phase of this project focused on understanding the core challenges faced by long-form writers and designing a system capable of addressing those limitations. Through an analysis of existing tools and their gaps, we identified the need for a context-aware, memory-driven writing assistant. Based on these insights, we finalized the complete system architecture, data models, modules, and design decisions required for CoWriteIA.

During FYP-I, we also implemented the foundational components of the system, including project indexing, semantic memory, and the initial version of the writing assistant. This phase successfully established the technical direction and groundwork necessary for building a fully functional, intelligent writing platform in the next development stage.

5.2 Future Work

The next phase will focus on completing and integrating all remaining modules into a cohesive system. Key areas include enhancing the context-aware writing agent, completing dialogue generation and character management features, and improving the gap analysis module. Additional work will involve research integration, optimizing semantic search performance, refining the user interface, and conducting extensive testing across all modules.

By the end of the next phase, the goal is to deliver a robust, polished, and reliable AI-assisted writing platform that supports writers throughout their creative workflow.

Bibliography

- [1] M. Abdin, J. Aneja, H. Awadallah, A. Awadalla, A. Awan, N. Bach, et al. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint*, 2024.
- [2] Grammarly. Features. Grammarly Website, n.d.
- [3] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. Realm: Retrieval-augmented language model pre-training. *Proceedings of the 37th International Conference on Machine Learning*, pages 3929–3938, 2020.
- [4] Omar Khattab and Matei Zaharia. Colbert: Efficient and effective passage search via contextualized late interaction over bert. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 39–48, 2020.
- [5] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.
- [6] Literature and Latte. Scrivener overview. Literature and Latte Website, n.d.
- [7] Notion. Everything you can do with notion ai. Notion Website, n.d.
- [8] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pages 3982–3992, 2019.

Appendix A

Appendices

A.1 Appendix A: System Diagrams

A.1.1 Use Case Diagram

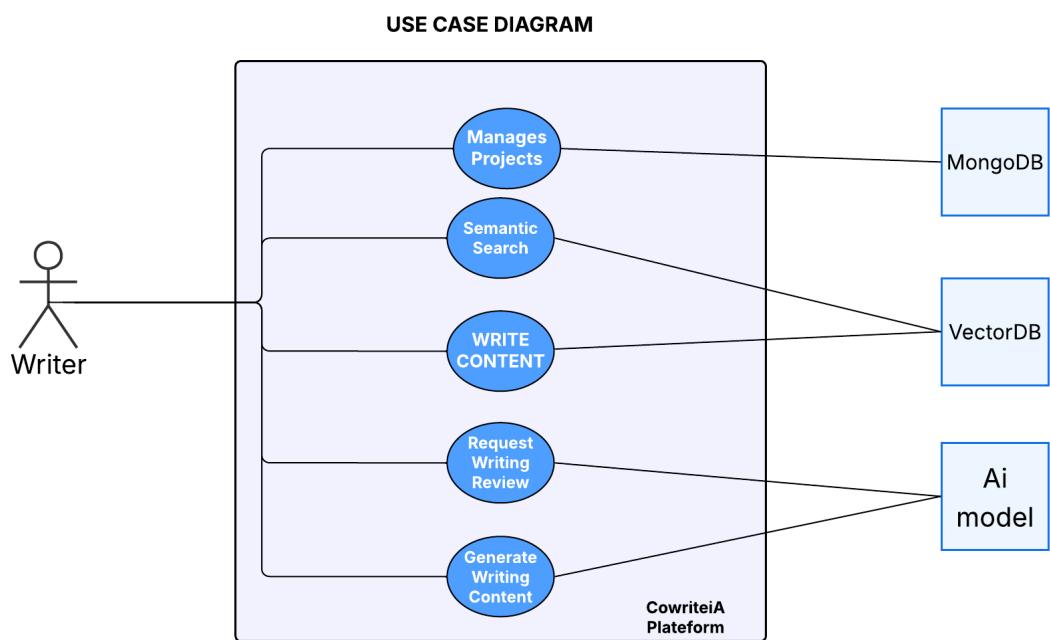


Figure A.1: Use Case Diagram CoWriteIA

A.1.2 Activity Diagram

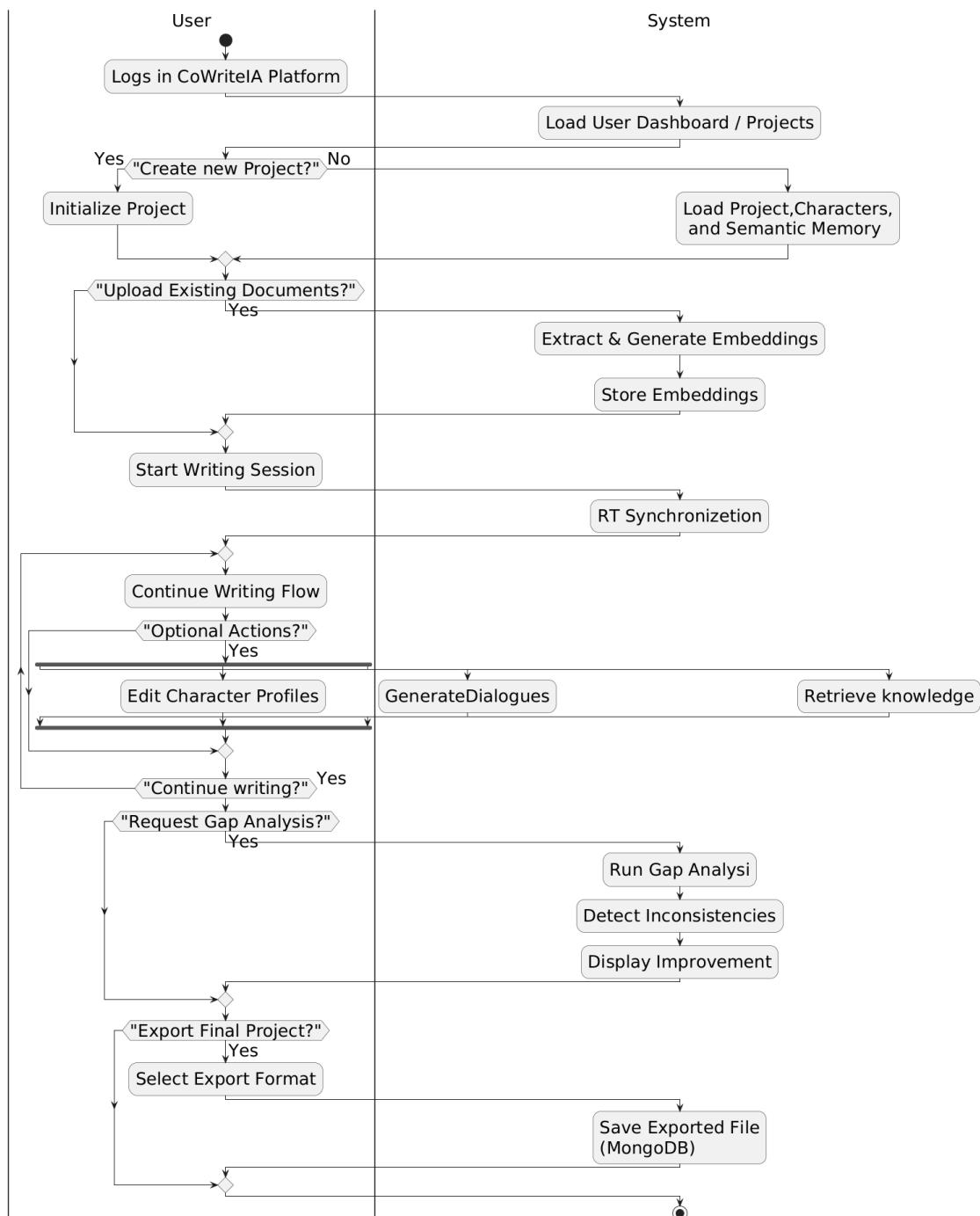


Figure A.2: Activity Diagram

A.2 Appendix B: Architecture Diagram

A.2.1 CoWriteIA Detailed Architecture Diagram

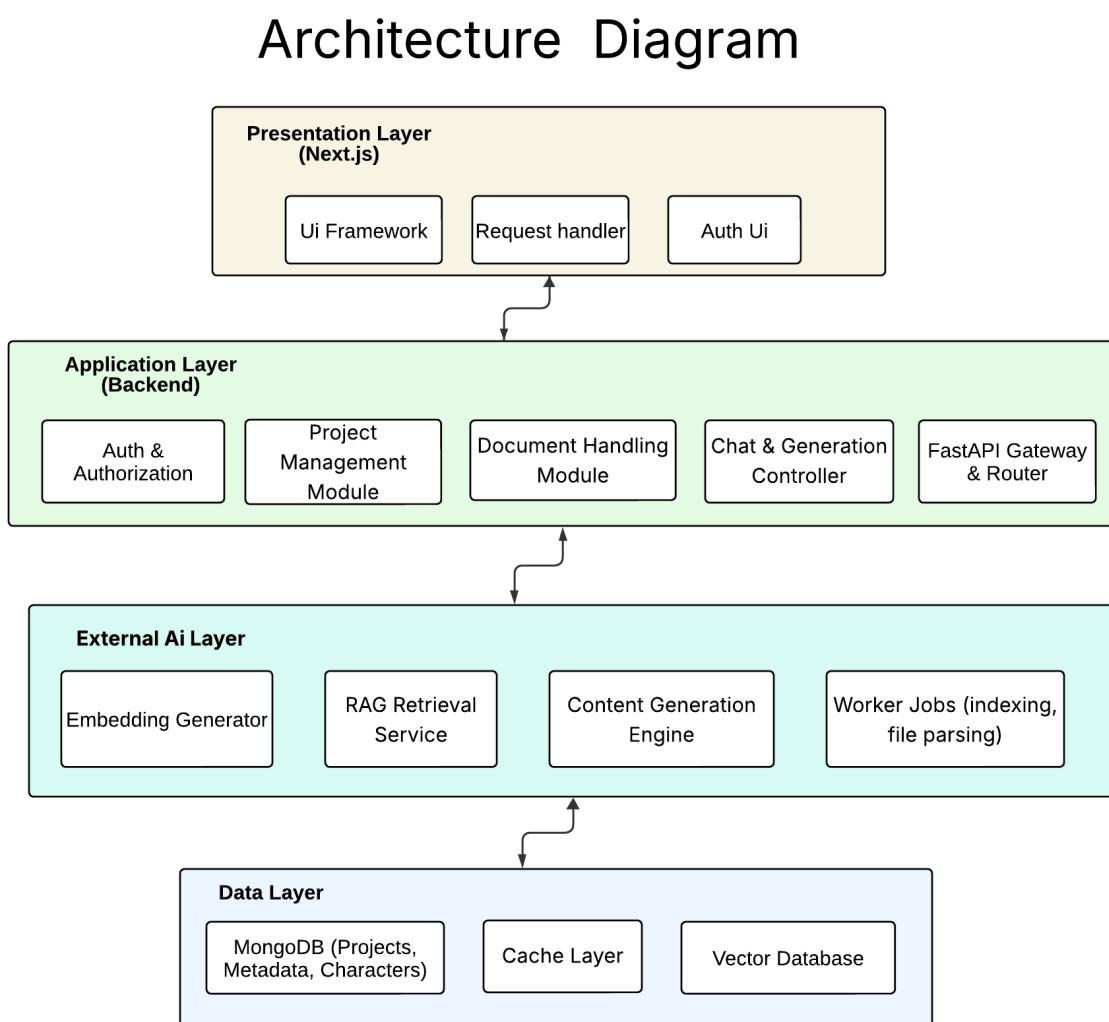


Figure A.3: CoWriteIA Multi-Tier Architecture Diagram

A.3 Appendix C: User Interface Screenshots

A.3.1 Dashboard Interface

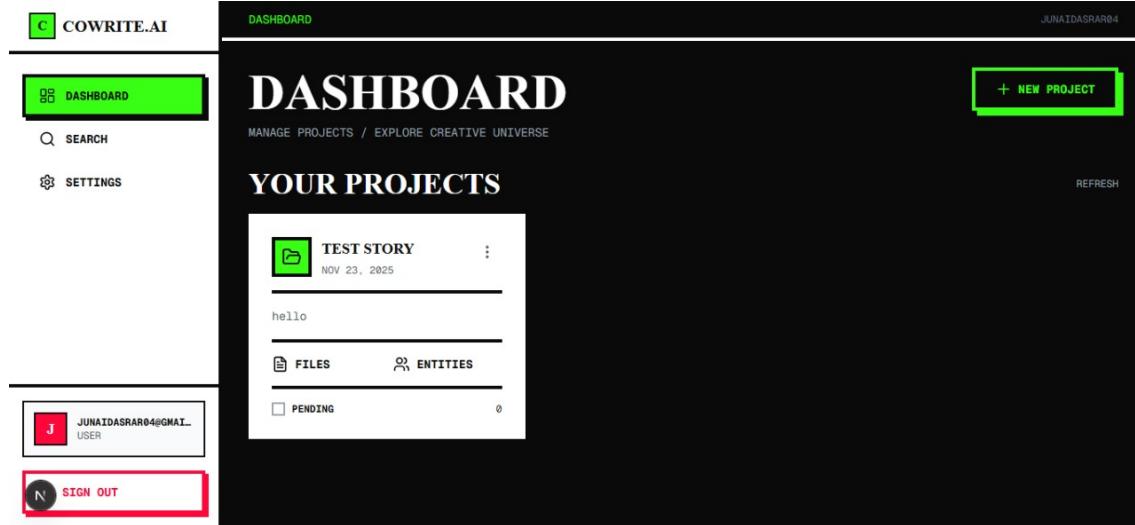


Figure A.4: CoWriteIA Dashboard Interface

A.3.2 Writing Environment Interface

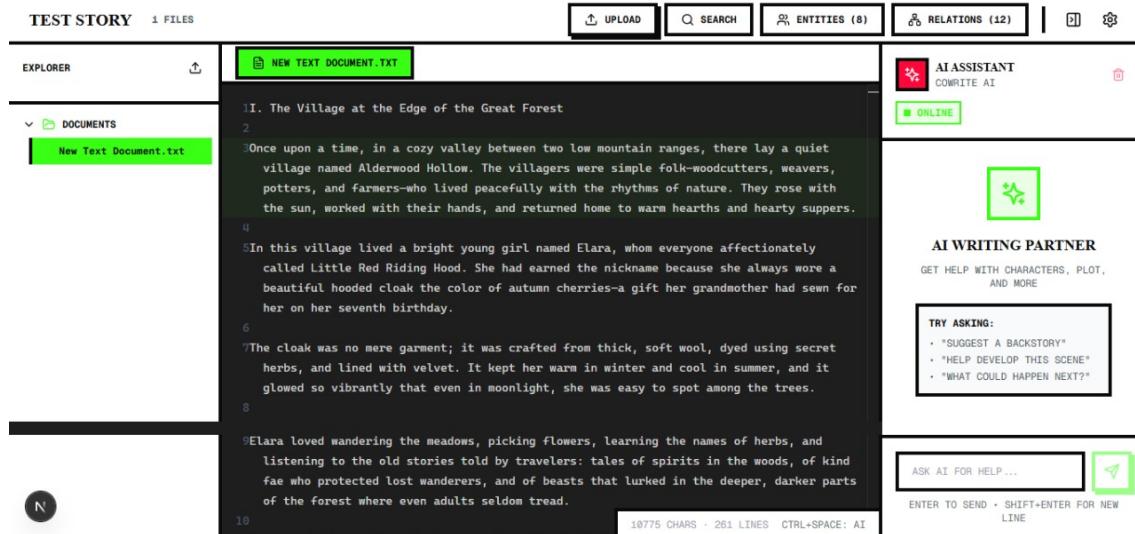


Figure A.5: Writing Environment Interface

A.3.3 Writing Assistant Interface

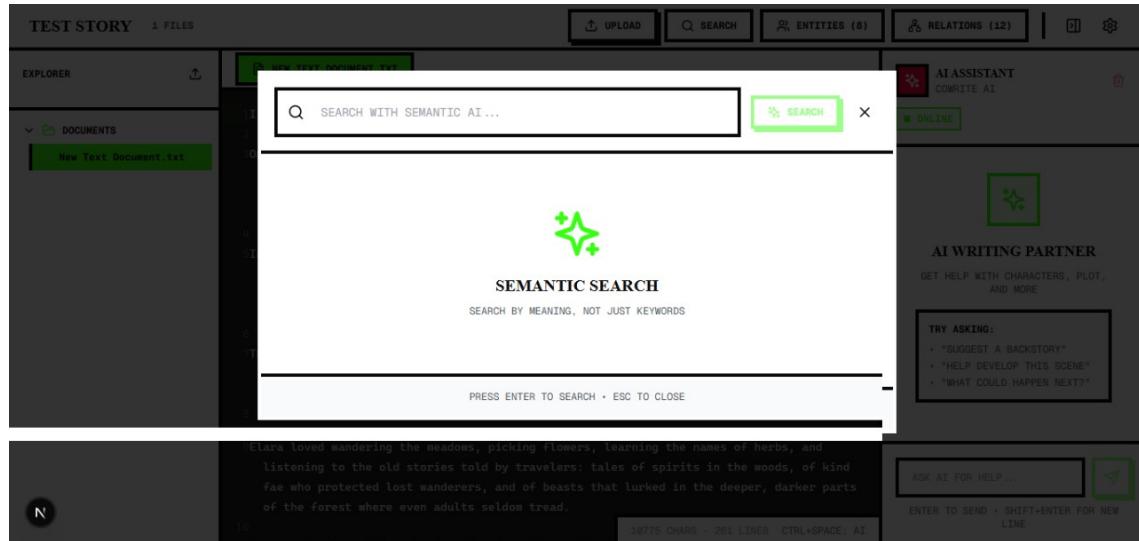


Figure A.6: Writing Assistant Interface

A.3.4 Login Interface

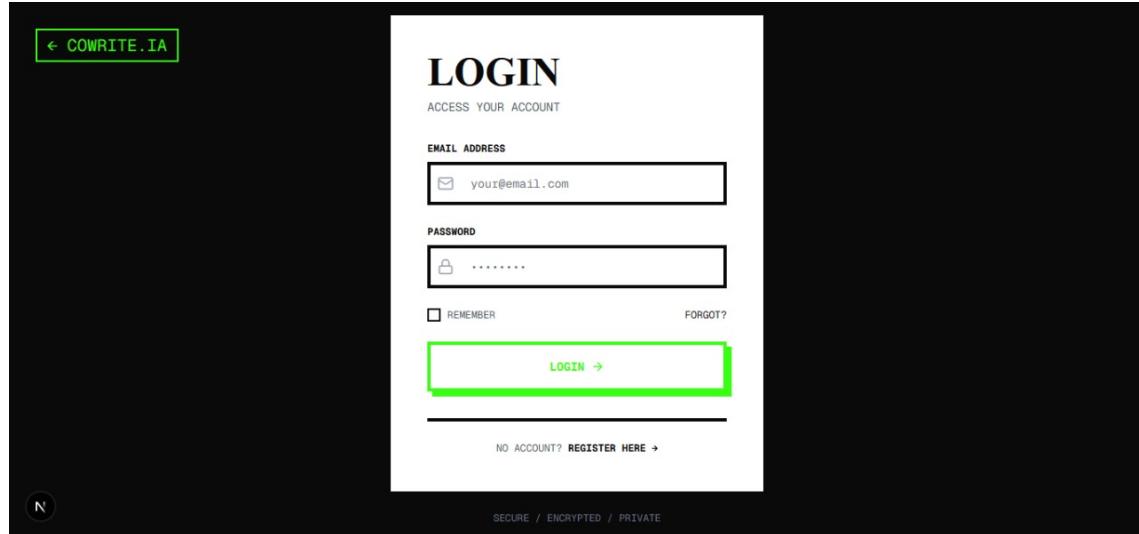


Figure A.7: Login Interface

A.4 Appendix D: Detailed Use Case Specification

A.4.1 UC-01: Generate Context-Aware Writing

Use Case ID	UC-01	
Use Case Name	Generate Context-Aware Writing	
Primary Actor	Writer (User)	
Secondary Actors	AI Processing Engine, Semantic Retrieval Module	
Preconditions	<ul style="list-style-type: none"> User is authenticated and logged into the system. A project exists with at least one uploaded or created document. Project content has been indexed and embeddings are stored. 	
Postconditions	<ul style="list-style-type: none"> AI-generated content is displayed to the user. Generated text may be saved to the project document. 	
Trigger	User selects the “Generate Content” option within the writing interface.	
Main Flow	<ol style="list-style-type: none"> User opens a project document in the writing environment. User requests content generation (e.g., continue writing, rewrite, expand). System retrieves relevant context using semantic search. System sends the prompt + context to the AI model. AI model generates new content. System displays the generated text to the user. User may choose to accept, edit, or regenerate the content. 	
Alternate Flow	AF-1 User requests generation with no specific prompt. <ul style="list-style-type: none"> (a) System uses recent document content as context. (b) AI generates a default continuation. AF-2 User modifies the context manually. <ul style="list-style-type: none"> (a) User provides custom text or notes. (b) System merges custom context with retrieved chunks. 	
Exceptions	E1 No embeddings found: System prompts user to index project files. E2 AI service unavailable: System displays fallback message and suggests retrying. E3 Empty or invalid user request: User is asked to refine or re-enter the prompt.	
Priority	High	36
Frequency of Use	Very frequent (core functionality)	

Table A.1: Detailed Use Case: Generate Context-Aware Writing