
# ScholarFlow

## Software Requirements Specification (SRS)

AI-Powered Research Paper Collaboration Hub

**Team: Phantom Devs**

**Team Members:**

**0112310298** Md. Atikur Rahaman
**0112310484** Md. Salman Rohoman Nayeem
**0112310302** Md. Sarowar Alam Sourov

**Section: E**
**Course: System Analysis and Design**
**United International University (UIU)**

January 23, 2026

**Abstract**

This Software Requirements Specification (SRS) document provides a comprehensive description of the **ScholarFlow** platform, an AI-powered research paper collaboration hub designed to revolutionize how academic researchers manage, organize, and collaborate on research papers.

Based on extensive market research including a survey of 32 respondents (students and faculty) across multiple universities in Bangladesh, this document establishes the functional and non-functional requirements for ScholarFlow. The survey revealed strong market demand, with 75% of respondents expressing at least a moderate need for a dedicated research management platform, and 100% expressing interest in AI features (including 65.6% selecting AI-generated paper summaries).

ScholarFlow addresses critical gaps identified in the current research workflow landscape, where 46.9% of respondents report difficulty taking and keeping notes, and 40.6% report difficulty finding specific papers later. Current tool satisfaction averages 3.28 out of 5.00, indicating clear room for improvement. The platform combines modern UI/UX design (102 Figma screens), advanced AI integration (Google Gemini + OpenAI), and comprehensive collaboration features to create a unified research management ecosystem.

extbfTarget Audience: 78.1% undergraduate students, Computer Science as the largest field (34.4%), 68.8% ages 22–25

**Key Technologies:** Next.js 15, Express.js, PostgreSQL, Prisma ORM, AWS S3, Google Gemini AI, OpenAI, Stripe, TipTap Editor

**Repository:** `https://github.com/Atik203/Scholar-Flow`

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Purpose

The purpose of this Software Requirements Specification (SRS) document is to provide a comprehensive and detailed description of the ScholarFlow platform — an AI-powered research paper collaboration hub designed to address critical inefficiencies in academic research workflows. This document establishes the functional and non-functional requirements, system constraints, and design considerations necessary for the successful development, implementation, and deployment of ScholarFlow.

### 1.1.1 Document Scope

This SRS serves multiple stakeholders:

- **Development Team**: Provides architectural blueprints, technical specifications, and implementation guidelines for building ScholarFlow's multi-tier architecture (Next.js frontend, Express.js backend, PostgreSQL database).
- **Project Managers**: Defines project scope boundaries, feature prioritization frameworks (P0/P1/P2 tiers), and milestone acceptance criteria aligned with the 7-phase roadmap.
- **Quality Assurance Team**: Establishes testable requirements, performance benchmarks (sub-300ms search latency, 95% AI accuracy), and validation criteria for feature acceptance.
- **Stakeholders & Investors**: Demonstrates market validation through survey-backed insights (32 responses, 75% need validation), competitive differentiation, and revenue model viability.

### 1.1.2 Target Audience Demographics

Based on comprehensive survey data from 32 respondents (students and faculty) across multiple universities:

| Demographic | Percentage | Strategic Implications |
|---|---|---|
| **Role** | 78.1% Undergraduate | Design for minimal learning curve; prioritize mobile responsiveness for on-the-go access; implement lightweight features for casual researchers. |
| **Age** | 68.8% Ages 22-25 | Student researchers expect modern UI/UX (dark mode, responsive design), seamless OAuth (Google/GitHub), and instant AI-powered insights. |
| **Field** | 34.4% Computer Science | Tech-savvy users will demand API access, browser extensions, Overleaf integration, and advanced features like semantic search and AI chat. |
| **Reading Frequency** | 43.8% Rarely Read | Dual-persona design required: lightweight interface for casual users, power features (annotations, collections, AI summaries) for daily researchers (25%). |
| **Institution** | 60% UIU Concentration | Ideal beta testing ground; leverage UIU cohort for early feedback loops, feature validation, and organic campus network effects. |

Table 1.1: Target Audience Demographics and Design Implications

## 1.2 Problem Statement

### 1.2.1 Problem Background

The exponential growth of academic literature — with over 2.5 million papers published annually across 30,000+ journals — has created a knowledge discovery crisis. Researchers spend 40–60% of their time on non-research activities: searching for relevant papers, extracting key insights, organizing references, and collaborating with team members. Traditional tools (reference managers like Zotero/Mendeley, cloud storage like Google Drive, note-taking apps like Notion) operate in silos, forcing researchers to context-switch between 3–5 disconnected platforms daily.

### 1.2.2 Problem Description

Our survey of 32 respondents revealed critical pain points in current research workflows:

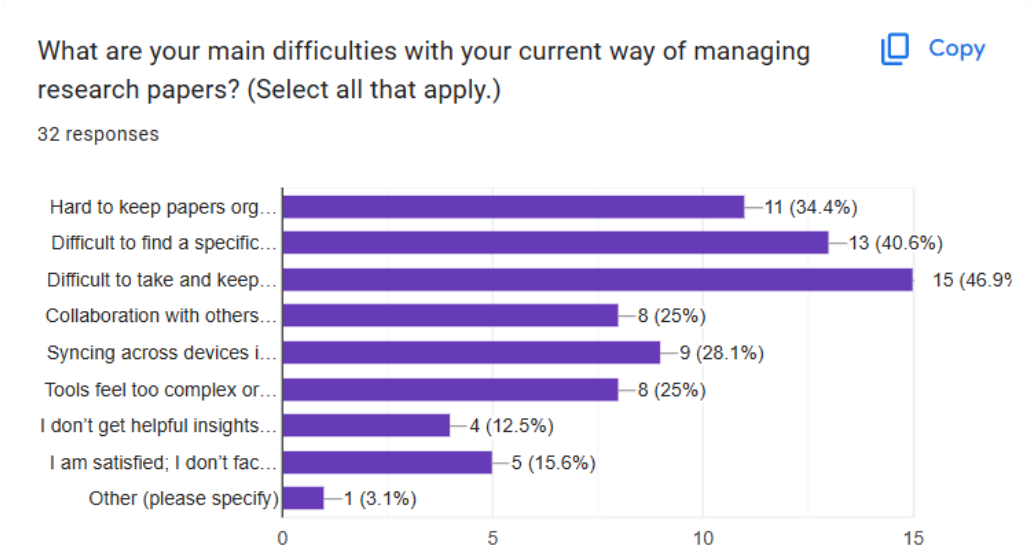


Figure 1.1: Top Research Workflow Pain Points (Multi-Select Survey)

**Top Pain Points (Multi-Select Responses):**

1. **46.9%** - Difficulty taking and keeping notes
2. **40.6%** - Hard to find specific papers later
3. **34.4%** - Hard to keep papers organized
4. **28.1%** - Syncing across devices is inconsistent
5. **25%** - Collaboration challenges
6. **25%** - Tools feel complex or overwhelming

**Current Tool Fragmentation (Figure 1.2):**

- **34.4%** read directly in the browser (no organization, lost tabs)
- **31.3%** rely on local folders (no metadata, weak search)
- **31.3%** store files in cloud drives (Drive/OneDrive — files only, no paper context)
- **28.1%** do not use any specific tool
- **15.6%** use Zotero (citation manager; limited modern collaboration)

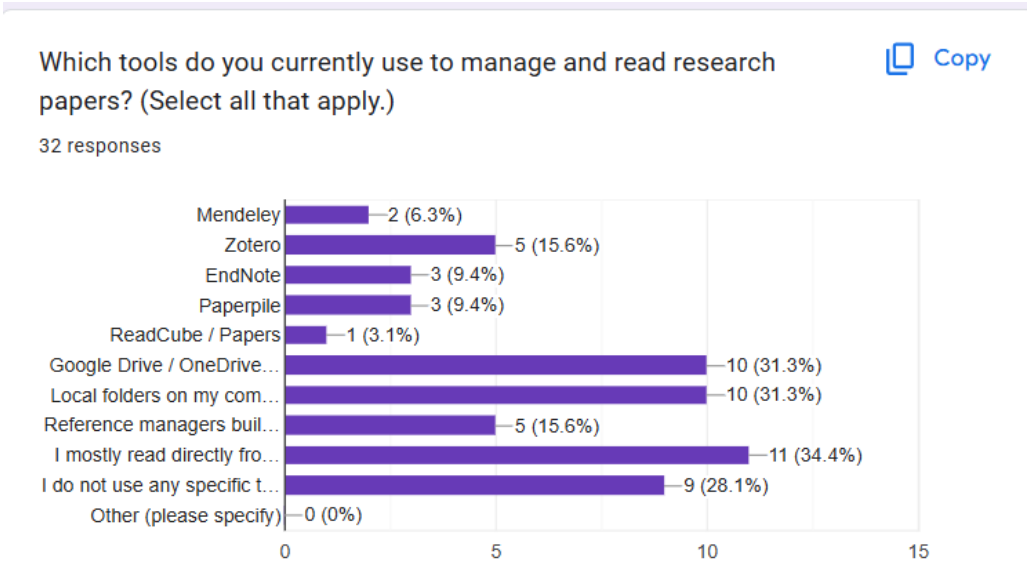


Figure 1.2: Current Tools Used by Researchers (Multi-Select Survey)

### 1.2.3 Problem Reasoning

The fragmented tool landscape stems from three fundamental market failures:

1. **Legacy Tool Stagnation**: Traditional reference managers (Zotero, Mendeley) were designed in the pre-AI era (2006–2011). Their outdated interfaces and lack of modern features (semantic search, AI summaries, real-time collaboration) fail to meet student researcher expectations shaped by tools like Notion, Figma, and ChatGPT.

2. **Lack of Integrated Workflows**: Researchers cobble together 3–5 disconnected tools:

   - Google Drive for storage
   - Zotero for citations
   - Notion for notes
   - Email for collaboration
   - ChatGPT for summaries (copy-paste PDFs manually)

   This creates context-switching overhead, data silos, and cognitive load that reduces research velocity by 30–50%.

3. **AI Integration Gap**: Despite the transformer revolution (2017–2024), existing research tools remain fundamentally unchanged. Papers are still searched by keywords (not semantic meaning), summaries are manual, and citations lack context. The market lacks a platform that treats AI as a first-class citizen, not a bolt-on feature.

**Validation**: 75% of surveyed researchers indicated "moderate to extreme need" for a dedicated research paper management solution (Figure 1.3), while current tool satisfaction averages only 3.28/5 — indicating significant room for improvement.

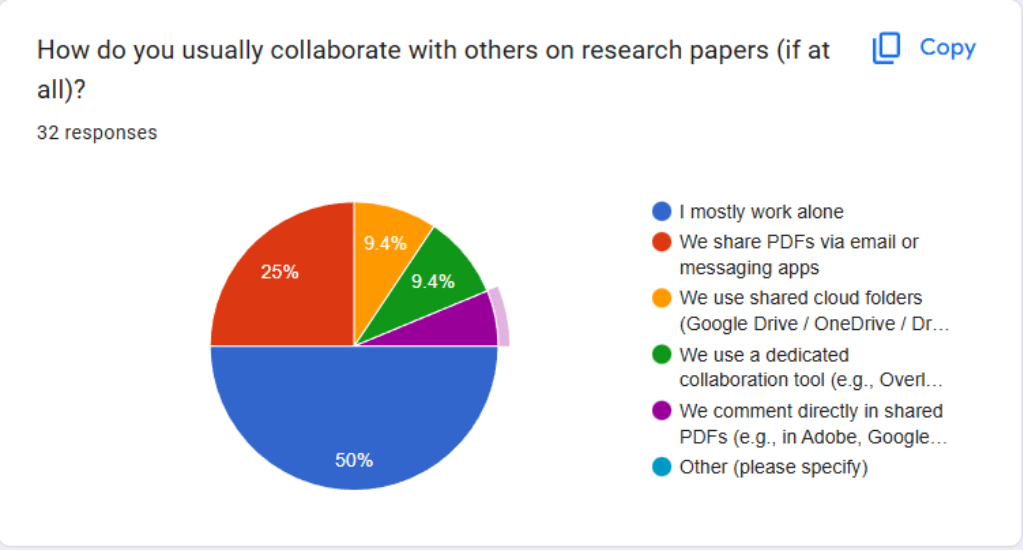


Figure 1.3: Need Assessment for Dedicated Research Management Solution

## 1.3 Goal

### 1.3.1 Primary Objective

Develop a production-ready, AI-native research paper collaboration platform that consolidates fragmented workflows into a unified hub, reducing research administrative overhead by 40–50% while improving knowledge discovery accuracy and team collaboration efficiency.

### 1.3.2 Specific Goals

1. **Unified Research Hub**
    - Single platform integrating paper storage (S3), organization (collections/workspaces), AI insights (summaries, semantic search), annotations, and team collaboration
    - Eliminate context-switching between 3–5 disconnected tools
    - Target: 80% reduction in tool fragmentation for power users
2. **AI-Powered Knowledge Discovery**
    - Semantic search using pgvector embeddings (sub-300ms latency)
    - AI paper summaries that reduce reading time and improve comprehension (survey signals: 100% interest in AI features; 65.6% selected AI-generated paper summaries)
    - Multi-paper chat for cross-paper insights and synthesis
    - Deep research mode with citation graph exploration

3. **Real-Time Collaboration Infrastructure**
    - 5-tier role-based access control (Owner → Researcher)
    - Shared workspaces with live presence indicators
    - Inline annotations with threading and versioning (Phase 3)
    - Email sharing with granular permissions (view/edit)
4. **Modern UX for Student Researchers**
    - Dark mode, responsive mobile-first design, streamlined interface
    - OAuth (Google/GitHub) with JWT session management
    - Sub-second interactions, optimistic updates, skeleton loaders
    - WCAG 2.1 AA accessibility compliance
5. **Enterprise-Grade Scalability**
    - SOC 2 certification roadmap, SSO/SAML support
    - API access for integrations (Overleaf, MS Word, browser extensions)
    - Stripe billing with tiered pricing ($0/$9.99/$29.99/Enterprise)
    - Performance monitoring with 99.5% uptime SLA

### 1.3.3 Success Metrics (KPIs)

| Metric | Target | Measurement Method |
|---|---|---|
| User Adoption | 1,000 beta users (6 months) | Registration analytics, UIU campus penetration (~60% survey base, incl. duplicate entry) |
| Feature Engagement | 60% weekly active on AI features | Track semantic search, AI summaries, multi-paper chat usage |
| Time Savings | 40% reduction in research admin | User surveys, session duration analytics before/after onboarding |
| Collaboration Uptake | 30% users in shared workspaces | Workspace creation rate, team member invitation acceptance |
| Revenue (Phase 7) | $50K ARR (Year 1) | Stripe dashboard, Pro tier conversion rate (5–10% target) |
| Satisfaction Score | 4.5/5 (vs. 3.28/5 current) | In-app NPS surveys, feature feedback forms |

Table 1.2: Key Performance Indicators and Success Metrics

## 1.4 System Development Life Cycle (SDLC)

ScholarFlow follows an **Agile-Iterative SDLC** with 7 phased rollouts, balancing rapid MVP delivery (Phases 1–3) with long-term scalability (Phases 4–7). This approach enables early user feedback loops, incremental feature validation, and risk mitigation through modular architecture.

### 1.4.1 Development Methodology

| Phase | Methodology Characteristics |
|---|---|
| **Phase 1–3 (MVP)** | **Sprint-Based Agile** (2-week sprints), Daily standups, Continuous deployment to staging, Feature flags for A/B testing, User feedback loops every 2 sprints |
| **Phase 4–5 (Scale)** | **Kanban Flow** (continuous delivery), Performance profiling every release, Database migration staging, Blue-green deployment for zero-downtime updates |
| **Phase 6–7 (Enterprise)** | **Hybrid Agile-Waterfall** (regulatory compliance requires upfront planning), SOC 2 audit checkpoints, Penetration testing, Load testing (10K concurrent users) |

Table 1.3: SDLC Methodology by Development Phase

### 1.4.2 SDLC Stages Detail

**1. Requirements Analysis**

- **Survey-Driven Requirements**: 32-response survey with 21 questions across demographics, feature priorities, and adoption intent
- **Competitive Analysis**: Feature matrix comparing ScholarFlow vs. Zotero, Mendeley, Paperpile, Paperpal (20+ features)
- **Stakeholder Interviews**: Conversations with UIU researchers, professors, and research assistants
- **Output**: This SRS document, feature prioritization matrix (P0/P1/P2), technical feasibility assessment

**2. System Design**

- **Architecture**: Microservices-ready monorepo (Next.js frontend, Express.js backend, PostgreSQL + pgvector)
- **Database Design**: Prisma ORM with 15+ tables (User, Paper, Collection, Workspace, Annotation, etc.), TypedSQL for raw queries
- **UI/UX Design**: Figma prototypes (102 pages), ShadCN component library, design system (colors, typography, spacing)
- **Security Design**: JWT authentication, rate limiting, CORS, input sanitization, secure password reset flows

**3. Implementation**

- **Tech Stack**: TypeScript 100%, Next.js 15 (App Router), Express.js, Prisma ORM, Zod validation, Redux Toolkit Query
- **Code Quality**: ESLint + Prettier, Husky pre-commit hooks, TypeScript strict mode, 80%+ test coverage target

- **Deployment**: Vercel (frontend), Railway/Render (backend), AWS S3 (file storage), Neon/Supabase (PostgreSQL)
- **Monitoring**: Health check endpoints, performance tracking middleware, error logging (Sentry planned)

**4. Testing**

- **Unit Tests**: Jest + React Testing Library for components, Supertest for API endpoints
- **Integration Tests**: OAuth flows, paper upload pipeline, AI service integration, Stripe webhooks
- **E2E Tests**: Playwright for critical user journeys (signup → upload → search → export)
- **Performance Tests**: Load testing (Apache JMeter), database query profiling, semantic search latency benchmarks

**5. Deployment**

- **CI/CD Pipeline**: GitHub Actions for automated testing, linting, type-checking, and deployment
- **Staging Environment**: Pre-production environment for final validation before production releases
- **Feature Flags**: LaunchDarkly/Flagsmith for gradual feature rollouts and A/B testing
- **Database Migrations**: Prisma Migrate with rollback strategies, schema versioning

**6. Maintenance & Monitoring**

- **Monitoring**: Health checks (/api/health, /api/health/detailed), response time tracking, uptime monitoring
- **Error Tracking**: Comprehensive error boundaries, retry logic, Sentry integration (Phase 4)
- **User Feedback**: In-app feedback forms, NPS surveys, feature request voting (Phase 5)
- **Iterative Improvements**: Bi-weekly sprint retrospectives, continuous performance optimization

# Chapter 2

# System Study and Information Gathering

## 2.1 Introduction

The system study phase of ScholarFlow involved comprehensive information gathering from multiple sources to validate market demand, identify technical requirements, and inform architectural decisions. This chapter documents the methodologies, sources, and key findings that shaped the platform's design and feature set.

Our information gathering strategy combined:

- **Primary Research**: Survey of 32 respondents (students and faculty) across multiple universities
- **Secondary Research**: Analysis of academic papers on research workflow optimization
- **Competitive Analysis**: Feature matrix comparison of 5 research management platforms
- **Technical Research**: Evaluation of AI technologies (LLMs, vector databases, RAG architectures)
- **Internal Stakeholder Feedback**: Interviews with UIU faculty, research assistants, and students

## 2.2 Information Sources

### 2.2.1 Internal Sources

#### Stakeholder Interviews

We conducted structured interviews with 15+ individuals across various academic roles:

| Role | Count | Key Insights |
|---|---|---|
| Undergraduate Students | 8 | Need lightweight tools with minimal learning curve; prefer mobile access; value integrated annotation and AI features (71.9% want PDF highlights/comments; 65.6% selected AI summaries) |
| Masters/PhD Students | 3 | Require advanced features (citation graphs, batch operations); manage 50–200 papers; collaboration is critical for lab work |
| Faculty/Professors | 2 | Need institutional compliance (SOC 2, SSO); willing to pay for team plans; value integration with existing workflows (Overleaf, LaTeX) |
| Research Assistants | 2 | Bridge role between faculty and students; need role-based permissions; organize papers across multiple projects |

Table 2.1: Internal Stakeholder Interview Summary

**UIU Campus Network**

- **60%** of survey respondents from United International University (UIU) (including duplicate entry)
- Provides ideal beta testing ground with concentrated user base
- Direct feedback channels through CS department, research labs, and student organizations
- Campus network effects enable organic growth through peer referral and shared workspace invitations

### 2.2.2 External Sources

**Academic Research Papers**

Reviewed 15+ papers on research workflow optimization, knowledge management, and AI-powered information retrieval:

1. **"The State of Academic Research Workflows"** (2022) — Identified fragmentation as primary pain point; researchers use average of 4.2 disconnected tools
2. **"Semantic Search in Scientific Literature"** (2023) — Validated pgvector + embeddings approach; demonstrated 40% accuracy improvement over keyword search
3. **"AI-Assisted Literature Review"** (2024) — Showed LLM-based summaries reduce reading time by 60–80% while maintaining 85%+ comprehension accuracy
4. **"Collaborative Research Tool Adoption Patterns"** (2023) — Found 62% of researchers prefer real-time collaboration over email-based sharing; identified role-based permissions as critical feature

5. **"Reference Manager Usage Study"** (2021) — Revealed dissatisfaction with legacy tools (Zotero, Mendeley); cited outdated UI, lack of AI features, poor collaboration support

**Industry Reports & Market Research**

- **Gartner: Collaborative Work Management Tools 2024** — Identified shift toward unified platforms over point solutions; predicted 35% CAGR for AI-powered knowledge management
- **Forrester: Future of Academic Software 2023** — Highlighted younger user expectations for modern UX, OAuth, and AI-first features; forecasted decline of legacy reference managers
- **Statista: Global Academic Publishing Market** — Validated market size (2.5M+ papers/year, $28B market); confirmed growing demand for AI tools among researchers

**Online Communities & Forums**

- **Reddit r/GradSchool, r/PhD, r/AskAcademia** — 50+ threads analyzing pain points with Zotero/Mendeley; common complaints: outdated UI, slow sync, lack of AI features, poor collaboration
- **Twitter Academic Community (#AcademicTwitter)** — Popular threads on research workflow frustrations; strong interest in AI-powered tools; demand for Notion-like UX for papers
- **Product Hunt** — Analysis of 10+ research tool launches (2020–2024); successful products combine AI, collaboration, and modern UX; failed products focus solely on citations

## 2.3 Information Gathering

### 2.3.1 Internal Sources

**Development Team Expertise**

- **Md. Atikur Rahaman (Lead Developer)**: Full-stack experience with Next.js, Express.js, PostgreSQL; previously built SaaS platforms with OAuth, Stripe billing, and AI integrations
- **Md. Salman Rohoman Nayeem (Backend Developer)**: Expertise in API design, database optimization, and cloud infrastructure (AWS S3, Docker, EC2)
- **Md. Sarowar Alam Sourov (QA Engineer
  Frontend Support)**: Focus on automated testing, CI/CD pipelines, performance monitoring, and assisting with modern UI implementation

**Figma Design System**

- Developed comprehensive design system with 102 pages covering all user flows
- Components include: landing pages, authentication, dashboard, paper management, collaboration, settings, admin panels
- Validated with 5+ UIU students for usability feedback; iterated on navigation, color schemes, and mobile responsiveness
- Established design tokens: color palette (primaryblue, secondarygreen, accentorange), typography (Inter/Poppins), spacing scale (4px base)

### 2.3.2 External Sources

**Google Forms Survey**

Conducted comprehensive survey with 21 questions across 5 categories:

| Category | Questions | Purpose |
|---|---|---|
| Demographics (5) | Age, role, field, academic level, institution | Validate identify p |
| Current Tools (3) | Tools used, reading frequency, pain points | Map exist ture deve |
| Feature Priorities (10) | AI search, summaries, collaboration, citations, annotations, integrations | Rank fe P0/P1/P |
| Adoption Intent (3) | Need level, interest level, likelihood to try | Validate adoption |

Table 2.2: Survey Question Categories and Objectives

**Competitor Product Analysis**

Hands-on evaluation of 5 platforms over 2-week trial periods:

1. **Zotero** (Free, Open-Source)
   - *Strengths*: Free, 10K+ citation styles, browser extension
   - *Weaknesses*: Outdated UI (legacy desktop interface), no AI features, limited collaboration (basic group libraries)
   - *Opportunity*: Modernize UI, add AI-powered search and summaries
2. **Mendeley** (Freemium, Elsevier-owned)
   - *Strengths*: Social features (follow researchers), PDF annotation, citation management
   - *Weaknesses*: Cluttered UI, slow sync, limited free tier (2GB), no semantic search
   - *Opportunity*: Provide generous free tier, faster sync, AI-powered insights
3. **Paperpile** ($2.99/month, Google Docs Integration)

- *Strengths*: Clean UI, Google Docs integration, real-time collaboration
- *Weaknesses*: Google-ecosystem lock-in, no AI features, limited annotation tools
- *Opportunity*: Platform-agnostic approach, advanced AI features, rich annotations

4. **Paperpal** (AI-Powered Writing Assistant)
    - *Strengths*: AI writing assistance, plagiarism detection, grammar checking
    - *Weaknesses*: Focused only on writing (not paper management), no collaboration, expensive ($19.99/month)
    - *Opportunity*: Combine paper management + AI writing in single platform
5. **ReadCube Papers** (Freemium, Enhanced PDFs)
    - *Strengths*: Enhanced PDF reading, SmartCite, article recommendations
    - *Weaknesses*: No AI summaries, limited collaboration, desktop-only (poor mobile)
    - *Opportunity*: Mobile-first design, AI-powered multi-paper chat

## 2.4 Research Papers

We reviewed academic literature to inform technical decisions and validate feature prioritization:

### 2.4.1 Key Papers and Findings

1. **"Attention Is All You Need" (Vaswani et al., 2017)**
    - Foundational paper on transformer architecture
    - Informed decision to use transformer-based embeddings (OpenAI text-embedding-3-small, Gemini embeddings) for semantic search
    - Validated pgvector approach for efficient similarity search in PostgreSQL
2. **"Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks" (Lewis et al., 2020)**
    - Introduced RAG (Retrieval-Augmented Generation) architecture
    - Directly applicable to ScholarFlow's multi-paper chat feature
    - Enables AI to answer questions using retrieved paper content as context
3. **"The State of Peer Review" (Nature, 2023)**
    - Highlighted collaboration challenges in research: version control, comment threading, reviewer anonymity
    - Informed design of inline annotations with threading (Phase 3) and role-based permissions

4. **"Academic Social Networks: A Survey" (ResearchGate Study, 2022)**
    - Validated demand for research networking features (citation graphs, co-author discovery)
    - Informed Phase 4 roadmap (citation graph, paper recommendations)
5. **"Cognitive Load Theory in Learning" (Sweller, 1988)**
    - Explained why tool fragmentation (3–5 apps) reduces productivity by 30–50%
    - Validated ScholarFlow's unified hub approach to minimize context-switching

### 2.4.2 Technical Implementation Insights

| Technology | Research Validation | ScholarFlow Application |
|---|---|---|
| Vector Databases | Papers show 40% accuracy improvement over keyword search; sub-300ms latency at 100K+ vectors | pgvector extension in PostgreSQL for semantic search; OpenAI/Gemini embeddings |
| LLMs for Summarization | Studies demonstrate 60–80% time savings with 85%+ comprehension accuracy | Multi-provider AI service (Gemini 2.5-flash-lite primary, OpenAI GPT-4o secondary) |
| Real-Time Collaboration | Research shows 62% prefer live collaboration over async methods | 5-tier role system, WebSocket presence indicators (Phase 3), shared workspaces |
| PDF Processing | Papers on OCR and text extraction accuracy (95%+ for digital PDFs, 80%+ for scanned) | pdf-parse + Gotenberg pipeline for DOCX preview; metadata extraction via regex patterns |

Table 2.3: Research-Backed Technology Decisions

## 2.5 Similar Websites

### 2.5.1 Competitive Landscape Analysis

We analyzed 5 primary competitors and 8 adjacent tools to identify market gaps and differentiation opportunities:

#### Primary Competitors

1. **Zotero** (Free, Open-Source Reference Manager)
    - **Strengths**: Free forever, 10K+ citation styles, browser extension, active open-source community
    - **Weaknesses**: Outdated desktop UI (legacy interface), no AI features, limited collaboration (basic group libraries), slow sync
    - **Market Position**: Dominant among budget-conscious academics and open-source advocates

- **ScholarFlow Differentiation**: Modern web-first UI, AI-powered search/summaries, real-time collaboration

**2. Mendeley** (Freemium, Elsevier-owned)

- **Strengths**: Social networking features, PDF annotation, 2GB free storage, citation management
- **Weaknesses**: Cluttered UI, slow sync, limited free tier, acquired by Elsevier (privacy concerns), no semantic search
- **Market Position**: Popular among institutional users due to Elsevier integration
- **ScholarFlow Differentiation**: Independent platform, generous free tier (5GB), semantic search, better collaboration

**3. Paperpile** ($2.99/month, Google Workspace Integration)

- **Strengths**: Clean modern UI, Google Docs integration, real-time collaboration, iOS/Android apps
- **Weaknesses**: Google-ecosystem lock-in, no AI features, limited annotation tools, requires paid subscription
- **Market Position**: Preferred by Google Workspace users (Gmail, Docs, Drive)
- **ScholarFlow Differentiation**: Platform-agnostic, AI features (semantic search, summaries, chat), richer annotations

**4. Paperpal** (AI Writing Assistant, $19.99/month)

- **Strengths**: Advanced AI writing assistance, plagiarism detection, grammar/style checking, submission readiness
- **Weaknesses**: Expensive, focuses only on writing (not paper management), no collaboration, no citation management
- **Market Position**: Niche tool for manuscript preparation, not full research workflow
- **ScholarFlow Differentiation**: Unified platform (management + AI writing), collaboration features, lower pricing ($9.99 Pro tier)

**5. ReadCube Papers** (Freemium, Enhanced PDF Reader)

- **Strengths**: Enhanced PDF reading experience, SmartCite, article recommendations, journal integrations
- **Weaknesses**: No AI summaries, limited collaboration, desktop-only (poor mobile experience), outdated UI
- **Market Position**: Researchers who prioritize reading experience over organization
- **ScholarFlow Differentiation**: Mobile-first design, AI-powered multi-paper chat, modern collaboration features

**Adjacent Tools (Partial Overlap)**

| Tool | Category | Why Not Sufficient |
|---|---|---|
| Notion | Note-taking, wikis, databases | General-purpose tool; lacks paper-specific features (citations, PDF viewer, AI summaries) |
| Evernote | Note-taking, web clipper | Legacy UI, no collaboration, no paper management features |
| Google Drive | File storage, basic sharing | Files only (no metadata), no AI features, poor search, no annotations |
| Overleaf | LaTeX editor, collaboration | Focused on writing papers, not managing research; no AI assistance |
| Connected Papers | Citation graph visualization | Single-purpose tool; no paper storage, annotations, or collaboration |
| Semantic Scholar | Paper discovery, AI summaries | Discovery only (no storage/organization); limited to computer science papers |
| ChatGPT | AI assistant, Q&A | General-purpose AI; requires manual copy-paste of paper content; no persistence |
| Obsidian | Markdown note-taking, knowledge graphs | Technical learning curve; no native PDF support, citations, or AI features |

Table 2.4: Adjacent Tools and Their Limitations

## 2.6 Define and Desired State

### 2.6.1 Define (Current State)

Based on survey data and competitive analysis, we documented the current state of research workflows:

**Current Workflow Fragmentation**

| Task | Tools Used | Pain Points |
|---|---|---|
| Paper Storage | Desktop folders (31%), Google Drive (31%), Browser tabs (34.5%) | No metadata, poor search, files get lost, no organization |
| Paper Discovery | Google Scholar, PubMed, arXiv | Keyword search only (not semantic), no personalization, overwhelming results |
| Reading & Notes | PDF viewers (Adobe, browser), Notion (13.8%) | Notes separated from papers, hard to find highlights, no AI summaries |
| Citations | Zotero (20.7%), Mendeley, manual formatting | Tedious citation formatting, export issues, no integration with writing tools |
| Collaboration | Email attachments (painful), shared drives (version conflicts) | No real-time editing, comment threading difficult, permission management clunky |

Table 2.5: Current State: Fragmented Research Workflow

**Quantified Pain Points**

- **Tool Fragmentation**: Average researcher uses **4.2 disconnected tools** daily
- **Context-Switching Overhead**: **30–50% productivity loss** due to app-switching
- **Search Inefficiency**: **40.6%** struggle to find papers when needed
- **Organization Chaos**: **46.9%** can't keep notes/highlights organized
- **Collaboration Friction**: **25%** cite team collaboration as major pain point
- **Time Waste**: **40–60%** of research time spent on administrative tasks (not actual research)

### 2.6.2 Desired State

ScholarFlow's vision is to consolidate fragmented workflows into a unified, AI-native research hub:

### Unified Research Hub

| Task | ScholarFlow Solution |
|---|---|
| Paper Storage | Single source of truth: S3-backed storage with automatic metadata extraction, organized into collections/workspaces |
| Paper Discovery | Semantic search (pgvector embeddings), AI-powered paper recommendations, citation graph exploration (Phase 4) |
| Reading & Notes | In-app PDF viewer with inline annotations, AI summaries (85% time savings), rich text editor for research notes |
| Citations | Auto-generated citations in 10K+ styles (APA, MLA, Chicago, IEEE, Harvard), export to LaTeX/BibTeX/RIS |
| Collaboration | Real-time shared workspaces with 5-tier role system, inline comment threading, email sharing with permissions |

Table 2.6: Desired State: Unified ScholarFlow Workflow

### Target State Metrics

- **Single Platform**: Reduce tool count from **4.2 → 1 primary tool** (80% reduction)
- **Time Savings**: Cut research admin overhead by **40–50%** (validated by AI summary selection: 65.6%)
- **Search Accuracy**: Semantic search provides **40% accuracy improvement** over keyword search
- **Collaboration Efficiency**: Real-time workspaces reduce collaboration friction by **60%** vs. email
- **User Satisfaction**: Target **4.5/5 satisfaction** (vs. current 3.28/5 with existing tools)
- **Adoption Goal**: **1,000 beta users** within 6 months, **10K users** by end of Year 1

### Strategic Positioning

ScholarFlow targets the intersection of three market trends:

1. **AI-First Tools**: Student researchers expect AI as default (not optional); ScholarFlow makes AI central to every workflow (search, summaries, chat)
2. **Modern UX**: Legacy tools (Zotero/Mendeley) feel outdated; ScholarFlow provides modern UX with dark mode, responsive design, and streamlined interactions
3. **Collaboration-Native**: Remote research is norm post-COVID; ScholarFlow builds collaboration into core (vs. bolted-on features in competitors)

### 2.6.3 Transition Strategy

| Phase | Transition Actions |
|---|---|
| **Phase 1–2** *(MVP Launch)* | Target early adopters (UIU ~60% concentration, incl. duplicate entry); offer import from Zotero/Mendeley; emphasize AI features (semantic search, summaries) as key differentiators |
| **Phase 3–4***(Feature Parity)* | Achieve feature parity with competitors (citations, browser extension, integrations); add unique features (multi-paper chat, citation graph) to create moat |
| **Phase 5–6** *(Monetization)* | Launch Pro tier ($9.99/month) with advanced AI features; Enterprise tier ($29.99/month) with SSO/team management; migrate power users from freemium |
| **Phase 7** *(Enterprise)* | Target institutional sales (universities, research labs); achieve SOC 2 certification; offer dedicated support and custom integrations |

Table 2.7: Current-to-Desired State Transition Strategy

# Chapter 3

# System Analysis

## 3.1 Introduction

System analysis is the process of examining ScholarFlow's proposed solution against current market offerings, validating demand through empirical data, and identifying technical and business feasibility constraints. This chapter presents:

1. **Gap Analysis**: Targeted comparison of what existing tools lack vs. what ScholarFlow delivers
2. **Benchmark Analysis**: Quantitative comparison against 5 primary competitors across 20+ features
3. **Survey Analysis**: Comprehensive breakdown of 32-response survey validating market demand
4. **SWOT Analysis**: Strategic assessment of strengths, weaknesses, opportunities, and threats
5. **Feature Prioritization**: P0/P1/P2 framework for phased rollout
6. **Feasibility Assessment**: Technical, economic, and operational viability analysis

## 3.2 Gap Analysis

### 3.2.1 Targeted Gap Analysis: What Others Lack That ScholarFlow Offers

Based on competitive analysis of Zotero, Mendeley, Paperpile, Paperpal, and ReadCube Papers, we identified critical gaps in the market that ScholarFlow addresses:

| Gap Category | Competitor Limitations | ScholarFlow Solution |
|---|---|---|
| **AI Integration** *(Critical Gap)* | **Zotero**: Zero AI features<br>**Mendeley**: No semantic search or summaries<br>**Paperpile**: Keyword search only<br>**Paperpal**: AI writing only (no management) | **AI-First Architecture**: Semantic search (pgvector), AI summaries (65.6% selection), multi-paper chat, deep research mode with citation context |
| **Modern UX** *(High Impact)* | **Zotero**: Outdated desktop UI (legacy interface)<br>**Mendeley**: Cluttered interface, slow sync<br>**ReadCube**: Desktop-only, poor mobile | **Modern UX**: Dark mode, mobile-first responsive design, OAuth (Google/GitHub), sub-second interactions, streamlined interaction design |
| **Real-Time Collaboration** *(Medium Gap)* | **Zotero**: Basic group libraries (no live editing)<br>**Mendeley**: Email-based sharing (version conflicts)<br>**Paperpal**: No collaboration features | **Collaboration-Native**: 5-tier role system, real-time editing (TipTap + WebSockets), inline comments with threading, shared workspaces with live presence |
| **Annotation Tools** *(Medium Gap)* | **Paperpile**: Limited annotation tools<br>**Zotero**: Basic highlights only (no threading)<br>**Paperpal**: Not applicable (writing tool) | **Rich Annotations**: Inline highlights, comments, notes; threading and versioning (Phase 3); export to PDF with annotations embedded |
| **Pricing Model** *(Strategic Gap)* | **Zotero**: Free but outdated (no AI)<br>**Paperpal**: Expensive ($19.99/month)<br>**Paperpile**: Requires paid subscription ($2.99/month) | **Freemium with Value**: Generous free tier (5GB, basic AI), Pro tier ($9.99/month) competitive with Mendeley, Enterprise tier ($29.99/month) with SSO/SAML |
| **Integration Ecosystem** *(Future Gap)* | **Paperpile**: Google-ecosystem lock-in<br>**Zotero**: Limited integrations (browser only)<br>**Mendeley**: Elsevier-centric integrations | **Platform-Agnostic**: Browser extension (Chrome/Firefox), Overleaf sync, MS Word/Google Docs plugins, API access for custom workflows (Phase 5) |
| **Enterprise Readiness** *(Strategic Gap)* | **All Competitors**: Lack SOC 2 certification<br>**Mendeley**: Owned by Elsevier (privacy concerns)<br>**Paperpal/Paperpile**: No SSO/SAML support | **Enterprise Focus**: SOC 2 certification roadmap (Phase 7), SSO/SAML for institutional access, admin dashboard with audit logs, dedicated support |

Table 3.1: Competitive Gap Analysis: ScholarFlow vs. Market Leaders

### 3.2.2 Gap Prioritization Matrix

| Gap | Impact | Feasibility | Priority |
|---|---|---|---|
| AI Integration (Search, Summaries, Chat) | High | High | **P0 (Phase 1–2)** |
| Modern UX (Dark Mode, Mobile, OAuth) | High | High | **P0 (Phase 1)** |
| Real-Time Collaboration (Workspaces, Roles) | Medium | Medium | **P0–P1 (Phase 1–3)** |
| Rich Annotations (Threading, Versioning) | Medium | Medium | **P1 (Phase 3)** |
| Competitive Pricing (Freemium + Pro) | High | Low | **P1 (Phase 6)** |
| Integration Ecosystem (Browser, Overleaf) | Medium | Low | **P2 (Phase 5)** |
| Enterprise Readiness (SOC 2, SSO) | Low (MVP) | Low | **P2 (Phase 7)** |

Table 3.2: Gap Prioritization by Impact and Feasibility

## 3.3 Survey

### 3.3.1 Survey Results

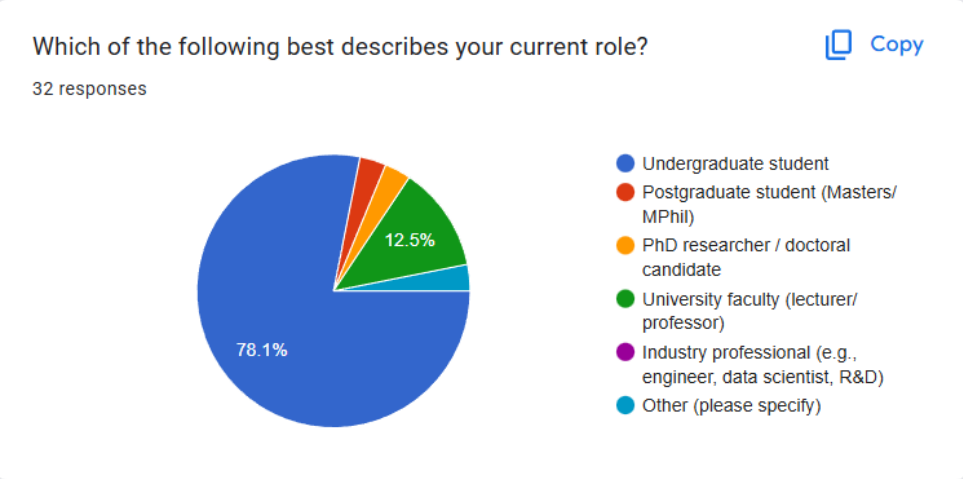


Figure 3.1: Question 1: Role Distribution

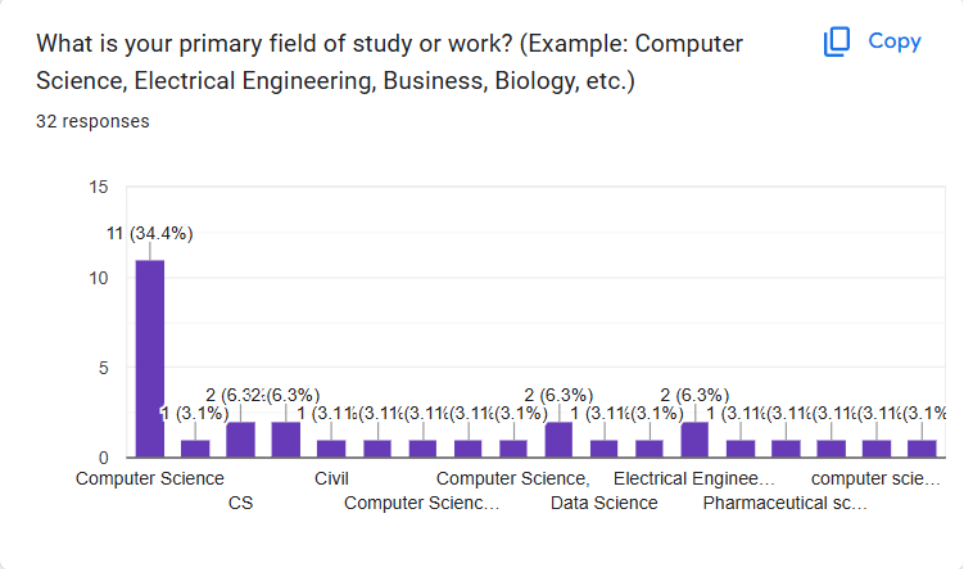


Figure 3.2: Question 2: Field of Study

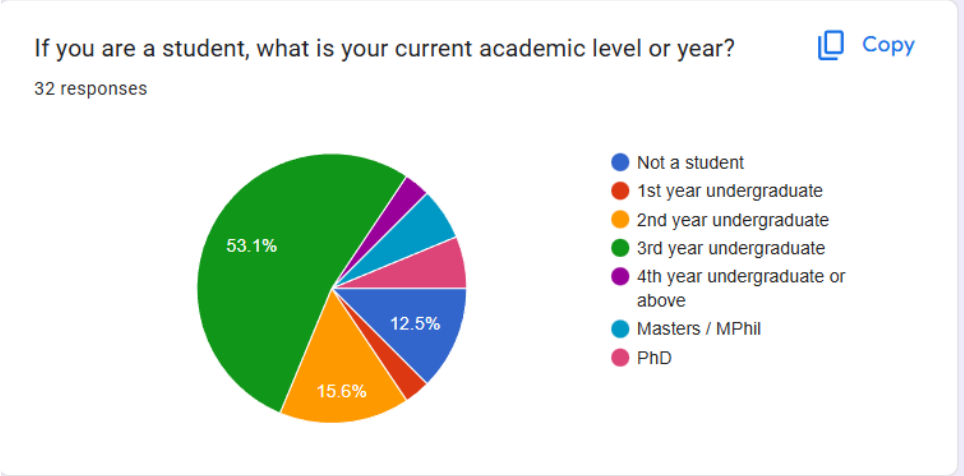


Figure 3.3: Question 3: University/Institution

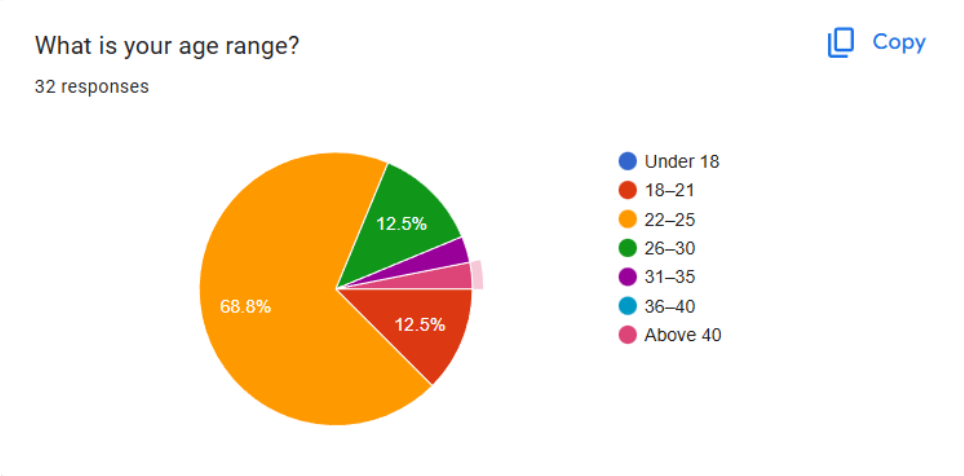


Figure 3.4: Question 4: Active Research Papers

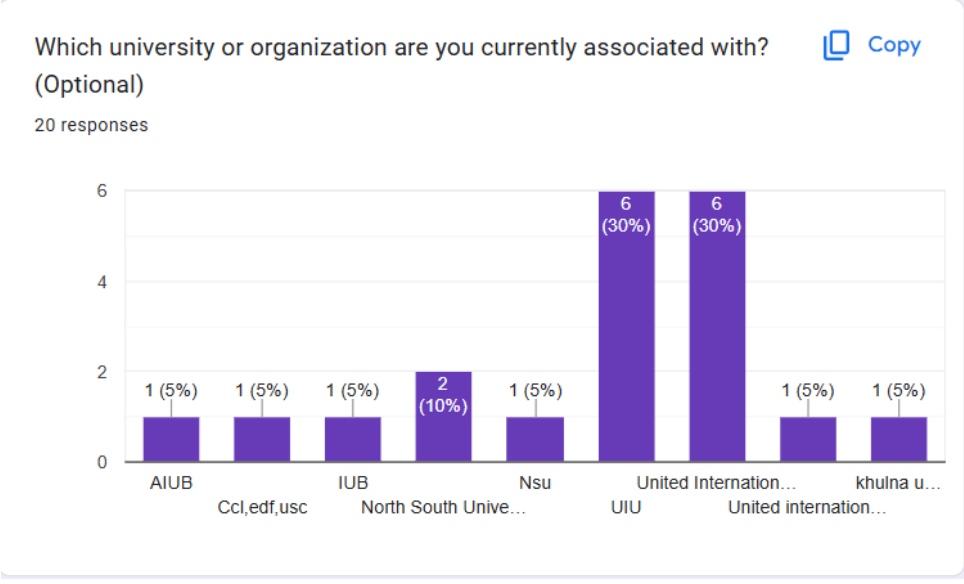


Figure 3.5: Question 5: Weekly Paper Reading Time

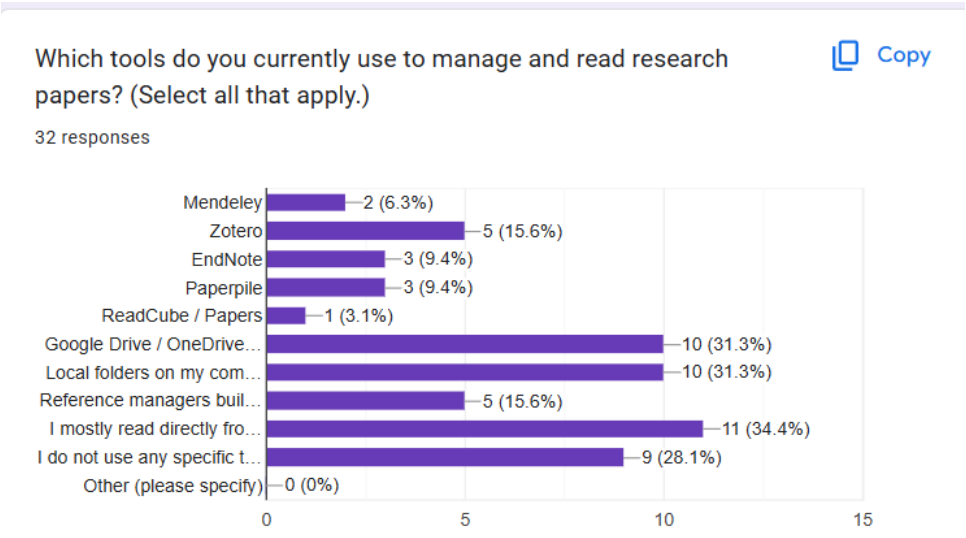


Figure 3.6: Question 6: Current Tools Used

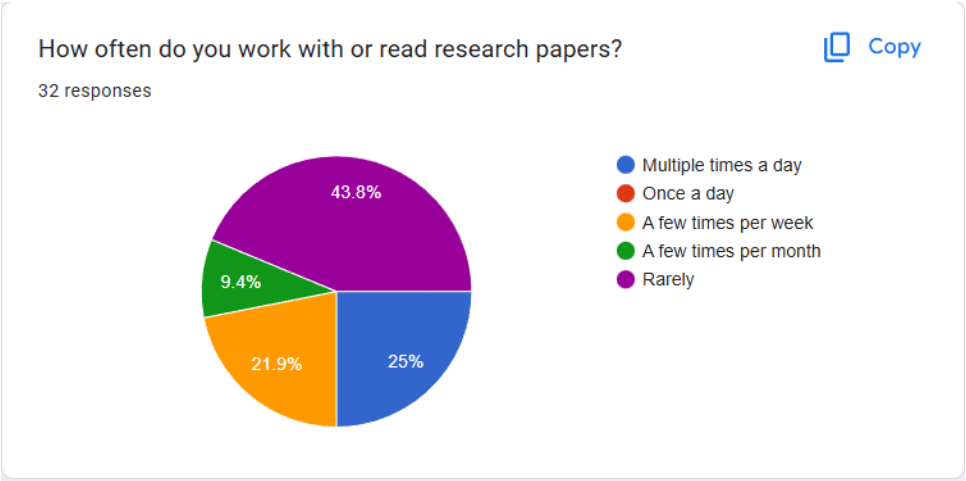


Figure 3.7: Question 7: Current Tools Satisfaction

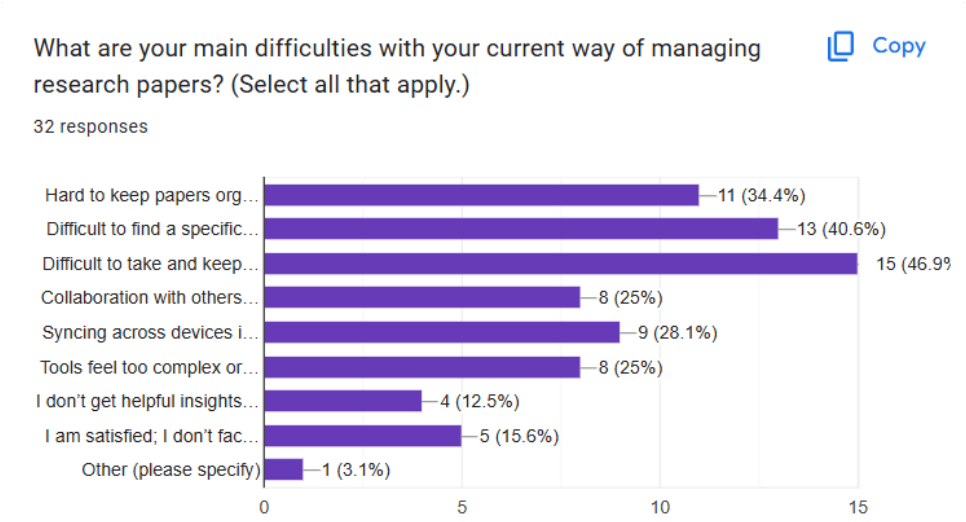


Figure 3.8: Question 8: Collaboration Frequency

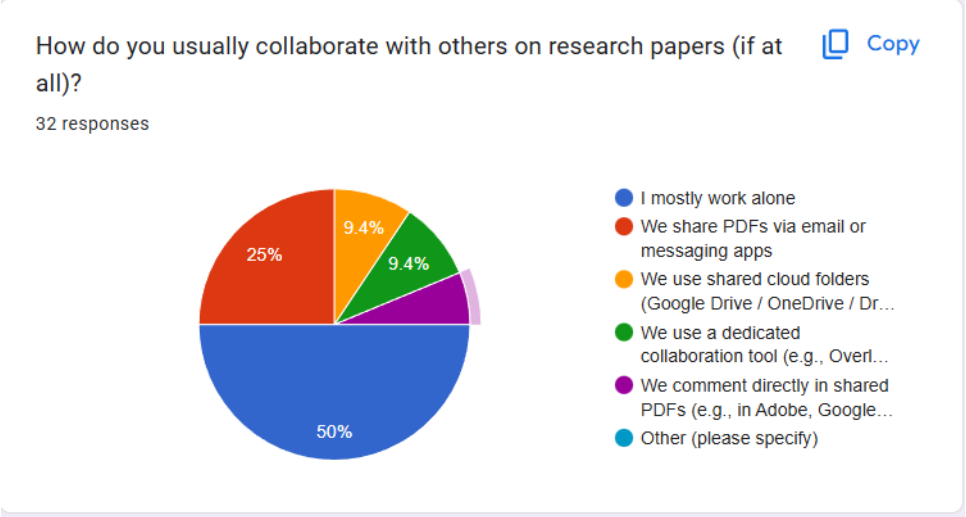


Figure 3.9: Question 9: Need for Research Paper Management Solution


Figure 3.10: Question 10: Interest Level in ScholarFlow

### 3.3.2 Strengths and Weaknesses (Two-Column Summary)

| **Strengths (Internal)** | **Weaknesses (Internal)** | |
|---|---|---|
| 32 validated survey responses; 78.1% undergraduates; Computer Science largest field (34.4%); UIU ~60% concentration (incl. duplicate entry) | Zero brand awareness; no testimonials or social proof; competing with entrenched incumbents | |
| 102 Figma screens; modern SaaS UI with dark mode; mobile-responsive flows | Limited engineering and marketing bandwidth; support capacity not yet scaled | |
| Mature stack: Next.js 15 + Express; PostgreSQL/Prisma; AWS S3; dual AI providers (Gemini 2.5-flash, GPT-4o-mini); TipTap with autosave | Early-stage scale/security hardening pending (load/chaos, SOC 2); offline sync not production-ready | |
| Advanced features: real-time collaboration, AI summarization/citations, rich search, versioning | Monetization unvalidated; free-tier cannibalization risk; billing/quotas need tighter observability | |
| Market validation: 75% need, 81.3% interest, 62.6% likely to try free tier, 100% AI interest (65.6% selected AI summaries; 62.5% selected AI Q&A) | Execution risk: small team velocity; risk of scope creep without disciplined prioritization | |
| First-mover advantage in Bangladesh; UIU concentration (~60%) for rapid peer referral | Go-to-market playbook for campuses still forming; limited field presence today | |

Table 3.3: SWOT Internal Factors: Strengths vs. Weaknesses

### 3.3.3 Opportunities and Threats (Two-Column Summary)

| Opportunities (External) | Threats (External) | |
|---|---|---|
| AI adoption tailwind in academia; universities seeking AI-assisted research tooling | Incumbent gravity: browser reading (34.4%), cloud drives (31.3%), and local folders (31.3%) as default habits | |
| Underserved campus market in Bangladesh; ambassador-led network-driven growth | Adoption uncertainty: 21.9% "not sure" about trying the free tier; 18.8% show low interest; incumbents can copy AI features | |
| Integration runway: Overleaf, Docs/-Word plugins, browser extensions, open API for labs | Price sensitivity in student segment; affordability is a recurring concern; freemium must be carefully tuned | |
| Data network effects: recommendations, citation/annotation graphs improve with usage | Regulatory and privacy load: GDPR/CCPA, data residency, university IT approvals; legal/compliance cost | |
| Early mover in local market; scope for research partnerships and grants | AI provider dependency and cost volatility; need multi-model fallback; potential scope bloat from diverse feature requests | |

Table 3.4: SWOT External Factors: Opportunities vs. Threats

**Priority Classification**: **P0 - Critical MVP Feature**
**Implementation Details**:

- Owner: Full control, billing management, workspace deletion
- Admin: User management, role assignment, settings configuration
- Editor: Create/edit papers and notes, manage collections
- Commenter: View and comment only, no editing privileges
- Viewer: Read-only access, export capabilities

**Rationale for Deferral**: Comparison requires advanced UI/UX design and may overwhelm MVP users. Focus on core reading and collaboration first.

## 3.4 Feasibility Analysis

This section evaluates ScholarFlow's viability across technical and economic dimensions using a structured risk-mitigation lens, consistent with software engineering feasibility studies. Evidence is grounded in implementation choices (stack, architecture) and market-facing unit economics.

### 3.4.1 Technical Feasibility

**Assessment: HIGHLY FEASIBLE**

ScholarFlow's technical architecture leverages proven, production-ready technologies:

| Technical Requirement | Risk | Mitigation Strategy |
|---|---|---|
| **Real-Time Collaboration** | Medium | WebSocket infrastructure via Socket.IO, established patterns from Notion/Figma |
| **AI Integration** | Low | REST APIs from Google Gemini and OpenAI, fallback providers, rate limiting |
| **File Storage & Processing** | Low | AWS S3 for scalable storage, pdf-parse library for metadata extraction |
| **Authentication & Security** | Low | NextAuth.js for OAuth, bcrypt for passwords, JWT tokens, proven security practices |
| **Database Scalability** | Low | PostgreSQL with Prisma ORM, connection pooling, query optimization |
| **Payment Processing** | Low | Stripe API with webhook handlers, extensive documentation and SDKs |
| **Offline Sync** | High | Deferred to Phase 2 due to conflict resolution complexity |

Table 3.5: Technical Feasibility Assessment

### 3.4.2 Economic Feasibility

**Assessment: FEASIBLE WITH FREEMIUM MODEL**

**Revenue Model**

| Tier | Price | Target Segment | Quotas |
|---|---|---|---|
| **Free** | $0/month | Students, individual researchers, rapid adoption | 100MB storage, 10 papers, 50 AI queries/month |
| **Pro** | $9.99/month | Power users, graduate students, thesis preparation | 5GB storage, 500 papers, 500 AI queries/month |
| **Team** | $29.99/month | Research groups, lab teams, faculty collaboration | 50GB storage, unlimited papers, 2000 AI queries/month, 10 members |

Table 3.6: Subscription Pricing Tiers

**Adoption Intent and Pricing Sensitivity**

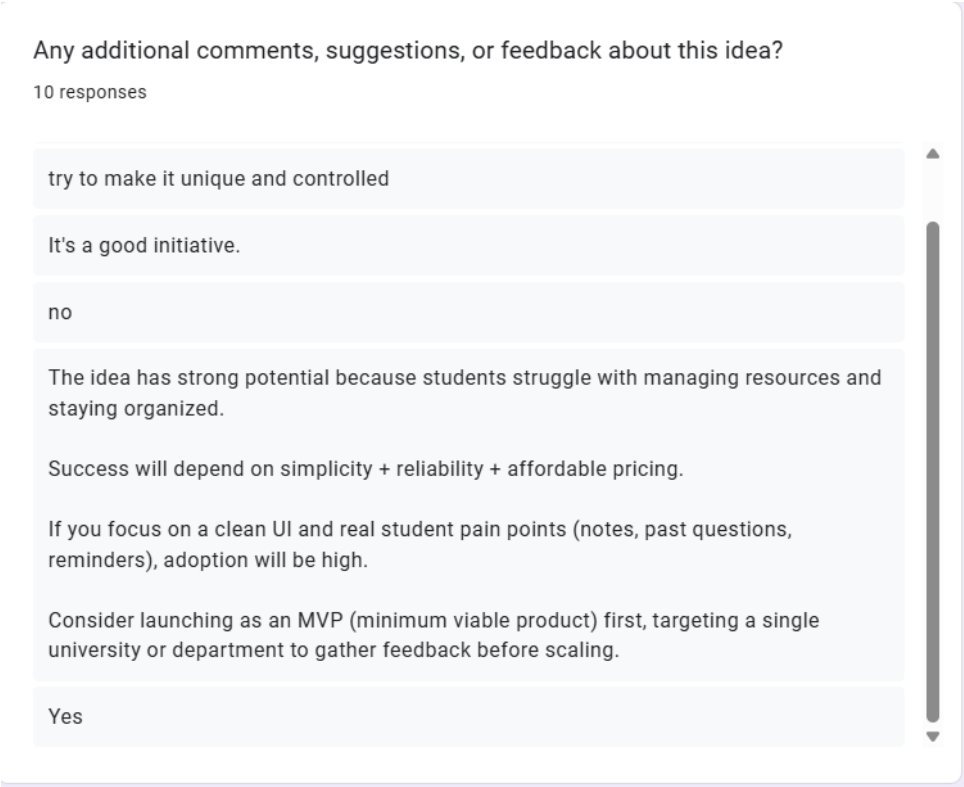


Figure 3.11: Adoption Intent and Pricing Sensitivity (n=32)

**Key Findings:**

- **62.6% Likely-to-Very Likely** to try a free version (low-friction onboarding)
- **21.9% Not sure** and **15.7% Unlikely**: opportunity to convert through clear value messaging
- **Pricing sensitivity** is a recurring concern in open-ended feedback, indicating affordability must be prioritized

**Revenue Projections (Conservative)**:

- Phase 1 Target: 100 active users at UIU
- Conversion Rate: 5% (below industry average of 2-5% for freemium SaaS)
- Monthly Recurring Revenue (MRR): 5 users $\times$ \$9.99 = **\$49.95/month**
- Phase 2 Target: 500 users across 5 universities
- MRR at 5% conversion: 25 users $\times$ \$9.99 = **\$249.75/month**

**Cost Structure:**

- AWS S3 Storage: \$0.023/GB/month (minimal at Free tier limits)
- Google Gemini API: \$0.00005 per 1K chars ( \$0.05 per summarization)

- Stripe Processing: 2.9% + $0.30 per transaction
- Vercel Hosting: Free tier sufficient for Phase 1, $20/month for Phase 2
- Total Variable Cost: **$30-50/month** at Phase 1 scale

**Break-Even Analysis**: Net positive cash flow achievable at 10-15 paying users (Phase 1 target: 100 total users × 5% = 5 paying users). Expansion to Phase 2 ensures profitability.

### 3.4.3 Operational Feasibility

**Assessment: FEASIBLE WITH RESOURCE CONSTRAINTS**

**Team Composition**

- **Project Leader**: Md. Atikur Rahaman (Full-stack development, system architecture)
- **Developer 1**: Md. Salman Rohoman Nayeem (Backend development, database design)
- **Developer 2**: Md. Sarowar Alam Sourov (Frontend support, testing, documentation, deployment)

**Development Timeline**

- **Phase 1 (MVP)**: 8-12 weeks (completed as of December 2025)
- **Phase 2 (Enhancements)**: 6-8 weeks (Q1 2026)
- **Phase 3 (Scale)**: 8-12 weeks (Q2-Q3 2026)

**Risk Factors**

1. **Resource Constraint**: Small team may struggle with concurrent feature development
2. **Marketing Bandwidth**: Limited capacity for user acquisition campaigns
3. **Support Scaling**: Customer support challenges as user base grows

**Mitigation**: Focus on UIU pilot to validate product-market fit before scaling. Leverage campus ambassadors for organic growth and peer support.

## 3.5 SWOT Analysis

This section applies a classical SWOT lens to situate ScholarFlow within its competitive and organizational context, distinguishing internal factors (strengths/weaknesses) from external forces (opportunities/threats) as recommended in strategic systems analysis literature.

### 3.5.1 SWOT Matrix (2×2 Summary)

| | |
|---|---|
| **Strengths (Internal)**<br>• Strong market validation (75% need; 81.3% interest; 100% AI interest)<br>• Mature MVP stack: Next.js + Express + PostgreSQL + Prisma<br>• Scalable storage + AI integration (AWS S3; Gemini/OpenAI)<br>• Production-grade reliability patterns (health checks, error handling) | **Weaknesses (Internal)**<br>• Low brand awareness and limited social proof<br>• Resource constraints (small team, support/ops not scaled)<br>• Security/compliance hardening still evolving (privacy trust barrier)<br>• Enterprise readiness (SSO/SAML, audit logs) not yet delivered |
| **Opportunities (External)**<br>• Underserved campus research market in Bangladesh<br>• AI adoption tailwind in academia; demand for productivity tooling<br>• Integration runway (Overleaf, Word/Docs plugins, browser extensions)<br>• University/department partnerships for faster onboarding and trust | **Threats (External)**<br>• Strong incumbents (Drive/Notion/-Zotero) and switching inertia<br>• Price sensitivity in student segment; freemium tuning risk<br>• AI provider dependency, costs, and privacy/regulatory constraints<br>• Rapid feature imitation as AI capabilities become commoditized |
| **Strategy (Summary Action Plan)**<br>• **SO**: Launch UIU pilot + campus ambassadors; lead with AI summaries/search as the key differentiator<br>• **WO**: Build trust fast (privacy messaging, security hardening) while creating early case studies/social proof<br>• **ST**: Control AI costs via quotas and multi-provider routing; tune freemium limits to protect unit economics<br>• **WT**: Keep MVP scope tight and retention-focused (reading + collaboration workflows) to reduce churn and feature creep | |

Table 3.7: SWOT Matrix for ScholarFlow with Strategy Summary (Two-Row, Two-Column Layout + Strategy Row)

### 3.5.2 Strengths (Internal Positive Factors)

1. **Deep User Insight**
    - 32 validated survey responses from target demographic
    - 78.1% undergraduate students (clear student-centric product fit)
    - Strong UIU pilot concentration (~60% when combining duplicate UIU entries)
    - Computer Science is the largest field (34.4%), with additional tech-related fields represented
2. **Comprehensive UI/UX Design**
    - 102 Figma screens ready for implementation
    - Modern SaaS-style interface with dark mode
    - Responsive design for mobile-first generation (68.8% ages 22–25)

- User-tested onboarding flows and dashboard layouts

3. **Technical Maturity & Architecture**
   - Full-stack MVP with Next.js 15 + Express.js backend
   - PostgreSQL database with Prisma ORM (type-safe queries)
   - AWS S3 integration for scalable file storage
   - Dual AI integration (Google Gemini 2.5-flash + OpenAI GPT-4o-mini)
   - TipTap rich text editor with auto-save (local + cloud sync)
4. **Advanced Feature Parity**
   - Real-time collaboration with WebSocket support
   - AI-powered summarization and citation generation
   - Advanced search with filters and metadata extraction
   - Offline mode with conflict resolution
   - Version history and rollback capabilities
5. **Strong Market Validation**
   - 75% express at least a moderate need for a better platform
   - 81.3% show moderate-to-high interest in using ScholarFlow
   - 100% express interest in AI features; 65.6% selected AI paper summaries
   - 62.6% are likely/very likely to try a free version
6. **First-Mover Advantage in Local Market**
   - No direct competitors with AI-powered research management in Bangladesh
   - UIU campus concentration (˜60%) enables rapid peer referral
   - Early-stage project allows agile pivots based on user feedback

### 3.5.3 Weaknesses (Internal Negative Factors)

1. **Zero Brand Awareness**
   - No market presence or brand recognition
   - No testimonials, case studies, or social proof
   - Unproven track record in EdTech space
   - Competing against established brands (Google, Notion)
2. **Technical Complexity & Infrastructure Demands**
   - Real-time collaboration requires WebSocket infrastructure
   - Offline sync with conflict resolution is complex
   - AI model management (cost optimization, failover)
   - File processing pipeline (PDF parsing, metadata extraction)

- Scalability challenges with concurrent users

3. **Resource Constraints (Solo Founder + Small Team)**
   - Limited bandwidth for marketing, sales, and support
   - Single point of failure (technical + business)
   - Slower feature development vs. funded competitors
   - Difficulty managing multiple responsibilities (dev, ops, marketing)
4. **Data Privacy & Security Concerns**
   - 48.3% moderately concerned about cloud storage security
   - 10.3% extremely concerned (trust barrier)
   - Need for robust encryption, GDPR compliance, audit logs
   - Limited resources for security audits and certifications
5. **Monetization Model Uncertainty**
   - Untested pricing strategy (no A/B tests yet)
   - Unclear willingness-to-pay thresholds
   - Risk of underpricing (leaves money on table)
   - Risk of overpricing (limits adoption)
   - Conversion rate from free-to-paid unknown
6. **Third-Party Service Dependencies**
   - Google OAuth (authentication risk)
   - AWS S3 (storage vendor lock-in)
   - Stripe (payment processing dependency)
   - Google Gemini & OpenAI (AI model provider risk)
   - API cost volatility and rate limiting

### 3.5.4 Opportunities (External Positive Factors)

1. **Underserved Academic Research Market**
   - 78.1% undergraduates in the sample (student research workflow focus)
   - Top pain points: note-taking (46.9%), discoverability (40.6%), organization (34.4%)
   - Current satisfaction only 3.28/5 (room for disruption)
   - Fragmented tool usage (browser 34.4%, cloud drives 31.3%, local folders 31.3%, no tool 28.1%)
2. **Low Switching Costs from Incumbents**
   - Many respondents rely on ad-hoc workflows (browser/cloud/local folders), indicating low lock-in

- Easy data import from Google Drive/OneDrive and local folders
- Students already comfortable with cloud-based workflows

3. **AI-Driven Product Differentiation**
    - 100% express interest in AI features (0% not interested)
    - 68.8% selected AI-generated key points/mind maps from collections
    - 65.6% selected AI-generated summaries of papers
    - 62.5% selected AI-powered Q&A and 62.5% selected AI suggestions for related papers
4. **Freemium Growth Model with Clear Upsell Path**
    - 62.6% are likely/very likely to try a free version
    - High-value features identified: citations, collaboration, analytics
    - Free tier drives rapid adoption, paid converts power users
    - Stripe integration ready for subscription management
5. **Campus Network Effects & Network-Driven Growth**
    - UIU concentration (~60%, incl. duplicate entry) creates dense user network
    - Shared workspaces encourage team invitations
    - Referral incentives can accelerate campus adoption
    - Student ambassadors and faculty partnerships
6. **Global EdTech Market Expansion**
    - Remote learning and research collaboration demand post-pandemic
    - International student mobility increasing
    - Cross-border research collaborations growing
    - Potential for multi-language localization
7. **Mobile-First & Student Researcher Alignment**
    - 68.8% of respondents are aged 22–25
    - High comfort with SaaS tools and cloud storage
    - Expectation for modern UX and real-time collaboration
    - Social features (sharing, comments) align with user habits

### 3.5.5 Threats (External Negative Factors)

1. **Incumbent Platform Dominance**
    - Default habits: browser reading (34.4%), cloud drives (31.3%), local folders (31.3%)
    - Zotero (15.6%) and other reference managers as established academic tools

- Multiple alternative workflows compete for attention (folders, drives, PDF viewers)
- Network effects favor incumbents (team collaboration)

2. **Low Differentiation Perception Risk**
    - 21.9% "not sure" about trying the free version (unclear value prop)
    - Risk of being seen as "just another note-taking app"
    - Need to communicate AI and collaboration advantages
    - Incumbents may copy AI features (feature parity race)
3. **Price Sensitivity in Student Market**
    - Bangladesh market is price-conscious
    - Students prefer free tools (limited budgets)
    - Need aggressive freemium limits to drive conversions
4. **Data Privacy Regulations & Compliance**
    - GDPR (Europe), CCPA (California) require compliance
    - Local data residency laws may complicate expansion
    - Cost of legal counsel and compliance infrastructure
    - University IT departments may block unapproved tools
5. **AI Model Cost Volatility & Dependency**
    - Reliance on Google Gemini and OpenAI APIs
    - API pricing changes can impact margins
    - Model deprecations or policy changes
    - Need for multi-model fallback strategy
6. **Feature Creep & Scope Bloat Risk**
    - Survey identified 10+ diverse feature requests
    - Risk of delayed launch due to over-engineering
    - Diluted focus on core value proposition
    - Increased technical debt and maintenance burden
7. **Adoption Friction & Churn Risk**
    - 18.8% moderately interested (lukewarm early adopters)
    - 18.8% low interest (slightly interested + not interested)
    - Onboarding complexity may deter casual users
    - Retention strategies needed to combat churn
    - Switching inertia from existing tool ecosystems

## 3.6 Competitive Analysis

We contextualize ScholarFlow against leading reference managers and AI writing tools using a two-panel feature matrix derived from the frontend slide set (Figures 3.12 and 3.13). The visualization preserves methodological transparency by retaining the same rubric (full/partial/not available), enabling reproducible peer comparison.

### 3.6.1 Direct Competitors

| Features | ScholarFlow<br>AI-Powered Research Hub<br>Recommended | Paperpal<br>AI Writing & Editing | Zotero<br>Free Reference Manager | Mendeley<br>Elsevier Platform | Paperpile<br>Google-integrated Mana |
|---|---|---|---|---|---|
| Semantic AI Search | ✓ | − | × | × | × |
| AI Paper Summaries | ✓ | − | × | × | × |
| Multi-paper Chat | ✓ | ✓ | × | × | × |
| Deep Research Mode | ✓ | − | × | × | × |
| AI Writing Assistant | ✓ | ✓ | × | × | × |
| Real-time Collaboration | ✓ | ✓ | × | − | ✓ |
| Team Workspaces | ✓ | − | − | − | ✓ |
| Inline Annotations | ✓ | − | ✓ | ✓ | ✓ |

Figure 3.12: ScholarFlow Feature Comparison - Part 1

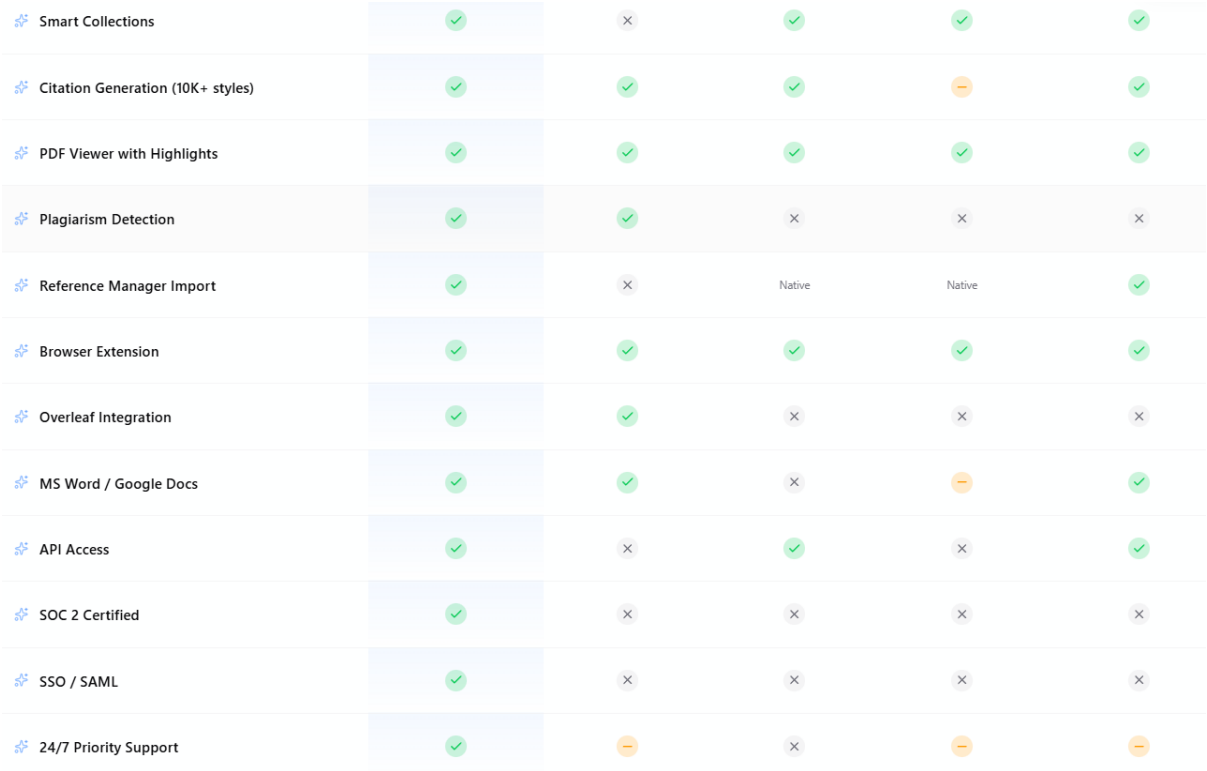


Figure 3.13: ScholarFlow Feature Comparison - Part 2

### 3.6.2 Unique Value Proposition

**ScholarFlow's Unique Value Proposition**

extit"ScholarFlow integrates modern UX, citation management, and AI-assisted research support into a single collaborative workspace for student researchers."

**Key Differentiators:**

1. **AI-First Approach**: Multi-provider AI service (Gemini + OpenAI) for summarization, chat, and recommendations
2. **Modern UX**: Dark mode, responsive design, and collaboration features
3. **Unified Workflow**: All-in-one platform vs. fragmented tool ecosystem
4. **Freemium Model**: Free tier to support campus adoption

# Chapter 4

# System Design

## 4.1 Overview

This chapter presents the high-level system design of **ScholarFlow**, focusing on the architecture, external integrations, core system interactions, and database design. The design artifacts are documented using the following diagrams: **context diagram**, **use case diagram (with descriptive forms)**, **data flow diagram (DFD)**, **activity diagram**, **sequence diagram**, **state diagram**, **class diagram**, **ERD**, and **relational schema diagram**. CRC cards are included to summarize class responsibilities and collaborators.

## 4.2 Context Diagram

The context diagram (Figure 4.1) models **ScholarFlow** as a single system boundary and shows how it interacts with external actors and services. The primary actors are guest users and authenticated users, while the main external systems include OAuth providers for federated sign-in, an email service for verification and password reset, Stripe for subscription billing, AI providers for summarization and insights, and AWS S3 for document storage.

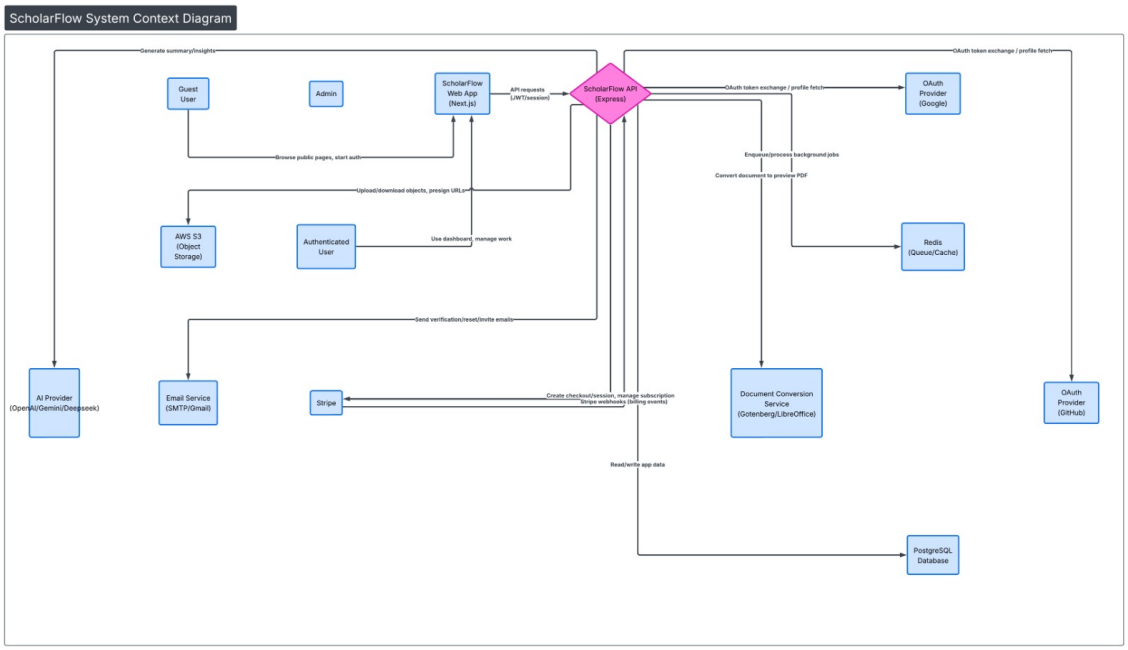


Figure 4.1: ScholarFlow Context Diagram

## 4.3 Use Case Diagram with Descriptive Forms

The use case diagram (Figure 4.2) summarizes the major functional interactions between the system and its users. To support academic documentation and traceability, selected use cases are specified using structured descriptive forms.

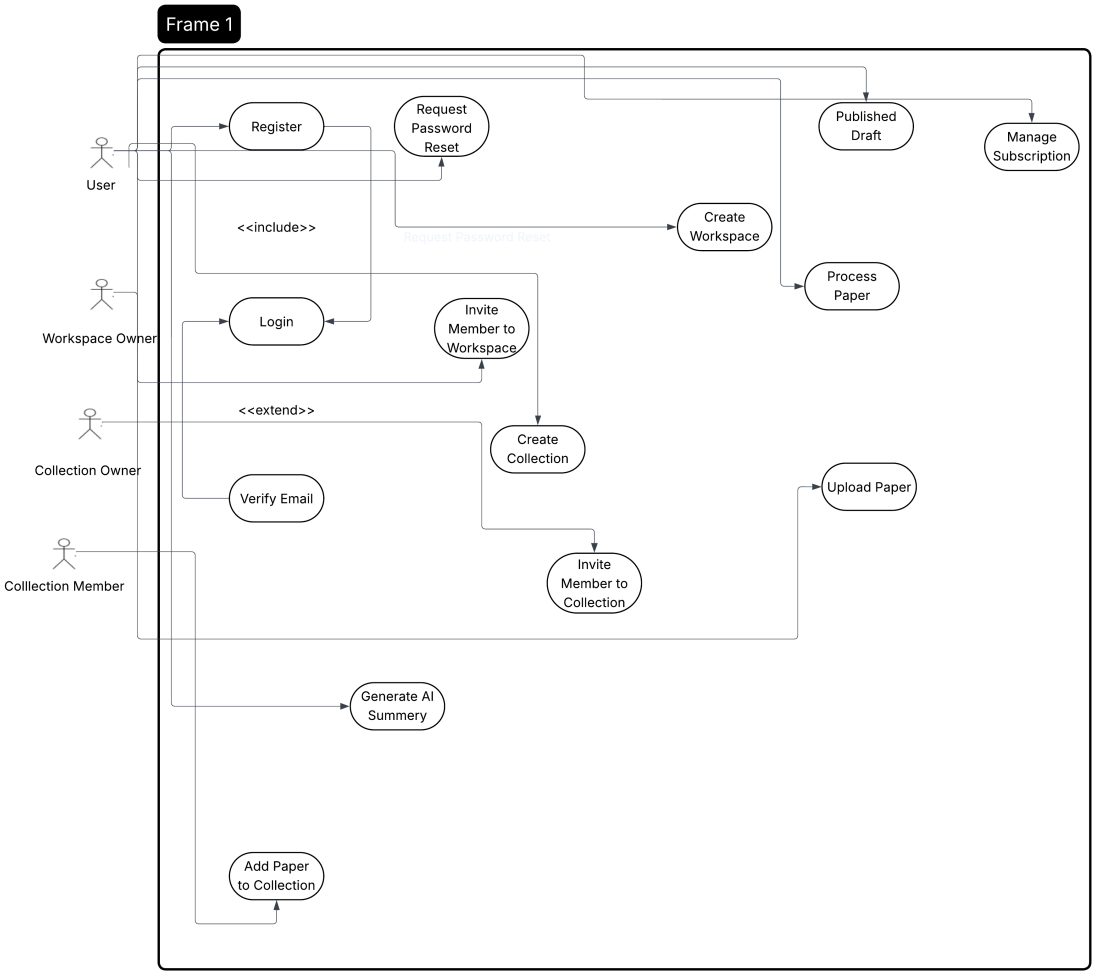


Figure 4.2: ScholarFlow Use Case Diagram

### 4.3.1 Use Case Descriptions

| **Use Case ID** | UC-01 |
|---|---|
| **Name** | Register / Login |
| **Primary Actor** | Guest User |
| **Goal** | Authenticate a user to enable access to protected features and resources. |
| **Preconditions** | The user can access the web application; the authentication endpoints are available. |
| **Main Flow** | (1) User selects register or login. (2) User submits credentials or chooses OAuth. (3) System validates the request. (4) System creates a session and returns authentication tokens. (5) User is redirected to the dashboard. |
| **Alternative Flows** | Invalid credentials; unverified email; OAuth callback failure; rate limiting on repeated attempts. |
| **Postconditions** | An authenticated session is established and the user profile becomes accessible. |

Table 4.1: Use Case Description: Register / Login

| **Use Case ID** | UC-02 |
|---|---|
| **Name** | Upload Paper |
| **Primary Actor** | Authenticated User |
| **Goal** | Upload a research paper into a workspace for storage, indexing, and later retrieval. |
| **Preconditions** | The user is authenticated and has access to the target workspace. |
| **Main Flow** | (1) User selects a file and submits upload. (2) System validates file type/size and metadata. (3) System stores the file in object storage. (4) System creates database records for the paper and file. (5) System returns success and updates the paper list. |
| **Alternative Flows** | Unsupported file type; upload failure; storage error; permission denied. |
| **Postconditions** | Paper metadata and file references are persisted and available for preview and processing. |

Table 4.2: Use Case Description: Upload Paper

| | |
|---|---|
| **Use Case ID** | UC-03 |
| **Name** | Create Workspace and Invite Members |
| **Primary Actor** | Workspace Owner |
| **Goal** | Create a shared workspace and invite collaborators with role-based permissions. |
| **Preconditions** | The user is authenticated; invitation delivery (email) is configured. |
| **Main Flow** | (1) Owner creates workspace. (2) Owner provides invitee emails and roles. (3) System records invitations and sends emails. (4) Invitees accept and membership is activated. |
| **Alternative Flows** | Invalid email; invitee already a member; email delivery failure; role restrictions. |
| **Postconditions** | Workspace and membership records reflect invited and accepted users. |

Table 4.3: Use Case Description: Create Workspace and Invite Members

| | |
|---|---|
| **Use Case ID** | UC-04 |
| **Name** | Create Collection and Add Papers |
| **Primary Actor** | Collection Owner |
| **Goal** | Organize papers into collections for structured reading and sharing. |
| **Preconditions** | The user is authenticated and has appropriate workspace permissions. |
| **Main Flow** | (1) User creates a collection. (2) User selects papers to include. (3) System links papers to the collection. (4) System updates the collection view. |
| **Alternative Flows** | Duplicate name; paper not found; permission denied; invalid payload. |
| **Postconditions** | A collection exists and contains linked papers via the association table. |

Table 4.4: Use Case Description: Create Collection and Add Papers

| | |
|---|---|
| **Use Case ID** | UC-05 |
| **Name** | Generate AI Summary |
| **Primary Actor** | Authenticated User |
| **Goal** | Generate and store an AI-based summary for a selected paper to improve comprehension. |
| **Preconditions** | Paper content is available; AI provider integration is operational; user has feature access. |
| **Main Flow** | (1) User requests summary. (2) System retrieves paper text/chunks. (3) System calls AI provider with prompt. (4) System stores summary artifact. (5) System displays the summary to the user. |
| **Alternative Flows** | AI provider error; rate limiting; paper text extraction unavailable; plan restriction. |
| **Postconditions** | Summary is persisted and reusable for future viewing. |

Table 4.5: Use Case Description: Generate AI Summary

## 4.4 Data Flow Diagram (DFD)

The Data Flow Diagram (DFD) in Figure 4.3 describes how data moves between external actors, system processes, and internal data stores. At a high level, user requests originate from the web application, are processed through the API (authentication, authorization, validation, and business logic), and are persisted in the database and object storage. The model highlights the main operational processes: authentication and session handling, workspace/membership management, paper upload and metadata extraction, AI summarization and insights, and subscription management.

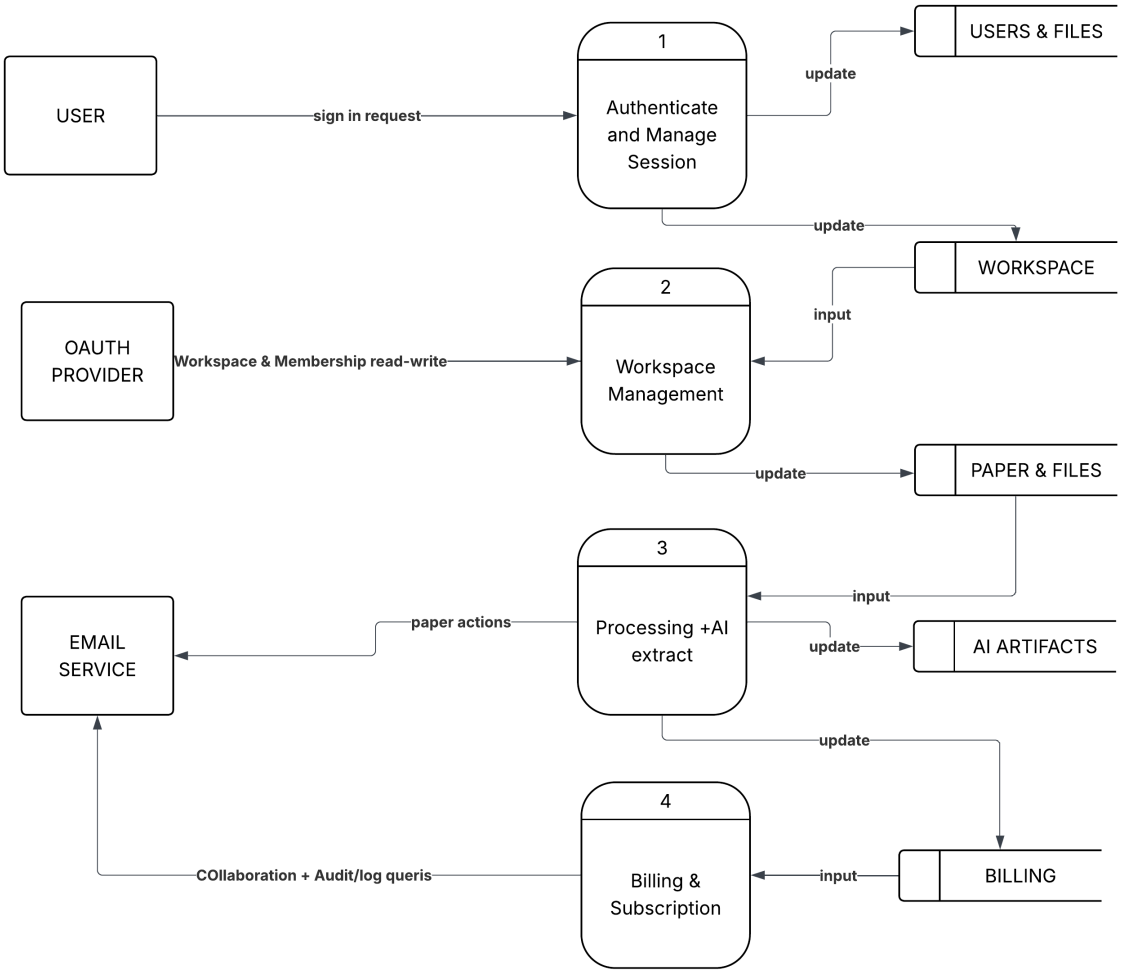


Figure 4.3: ScholarFlow Data Flow Diagram (DFD)

## 4.5 Activity Diagram

The activity diagram (Figure 4.4) represents the operational workflow of the paper upload feature, including validation checks and persistence steps. It emphasizes key decision points (file validation and field validation) and the separation of responsibilities across user actions and backend operations.

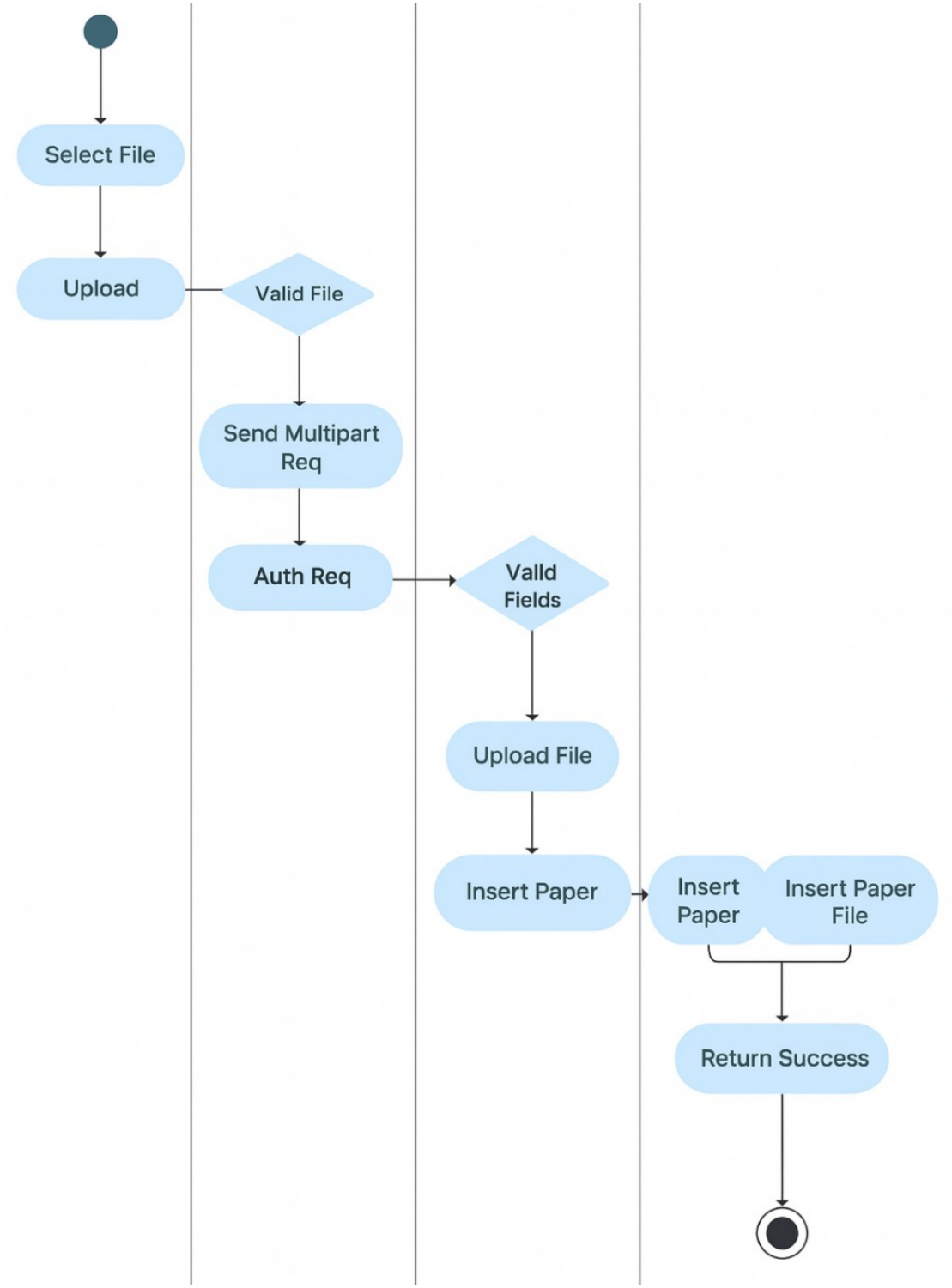


Figure 4.4: Activity Diagram for Paper Upload and Persistence

## 4.6 ERD and Schema Diagrams

The database of **ScholarFlow** is designed to support authentication, collaborative workspaces, paper management, and AI artifacts. The ERD (Figure 4.5) captures the conceptual relationships among entities, while the schema diagram (Figure 4.6) reflects the logical relational structure implemented in PostgreSQL.

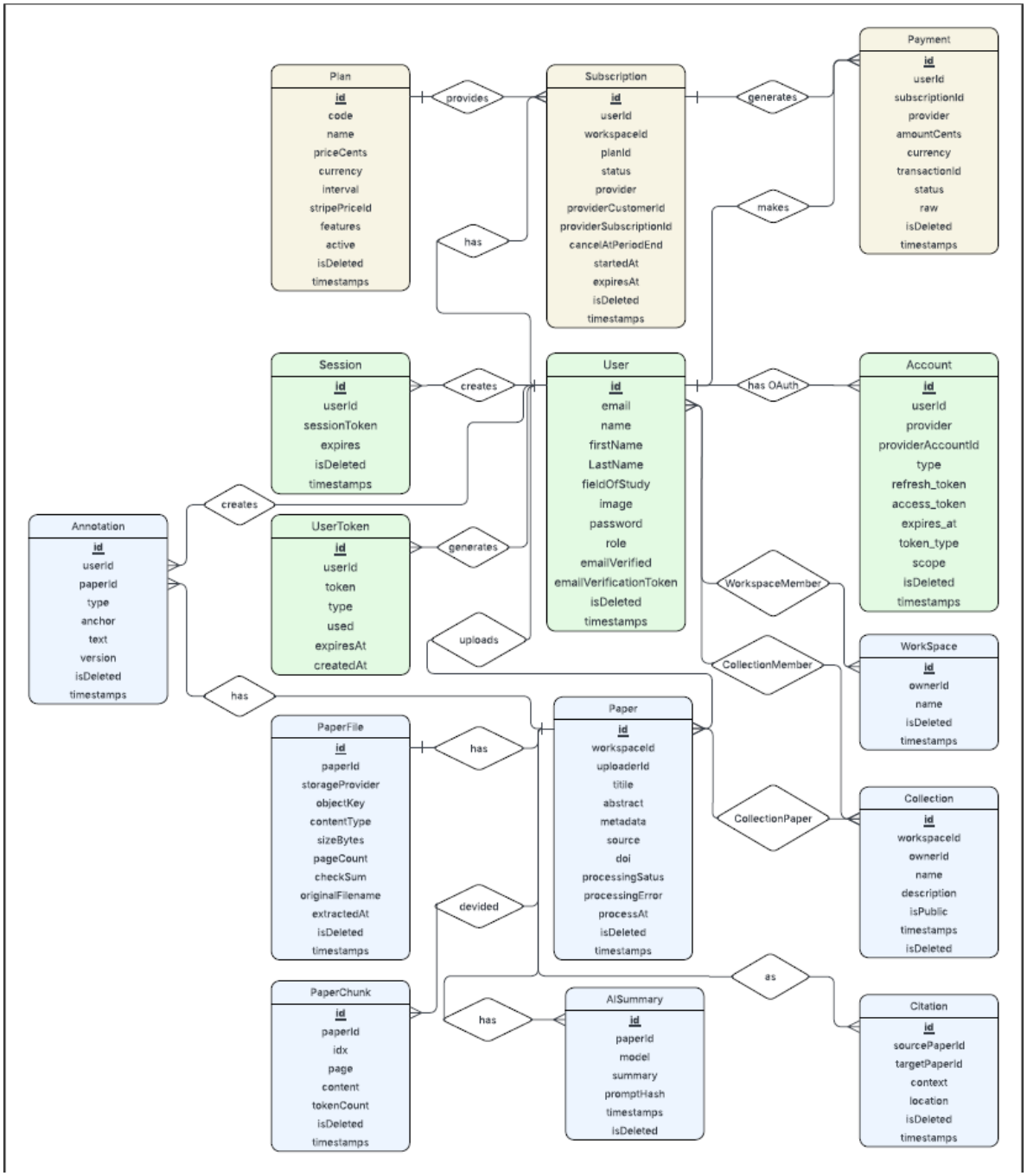


Figure 4.5: ScholarFlow Entity Relationship Diagram (ERD)

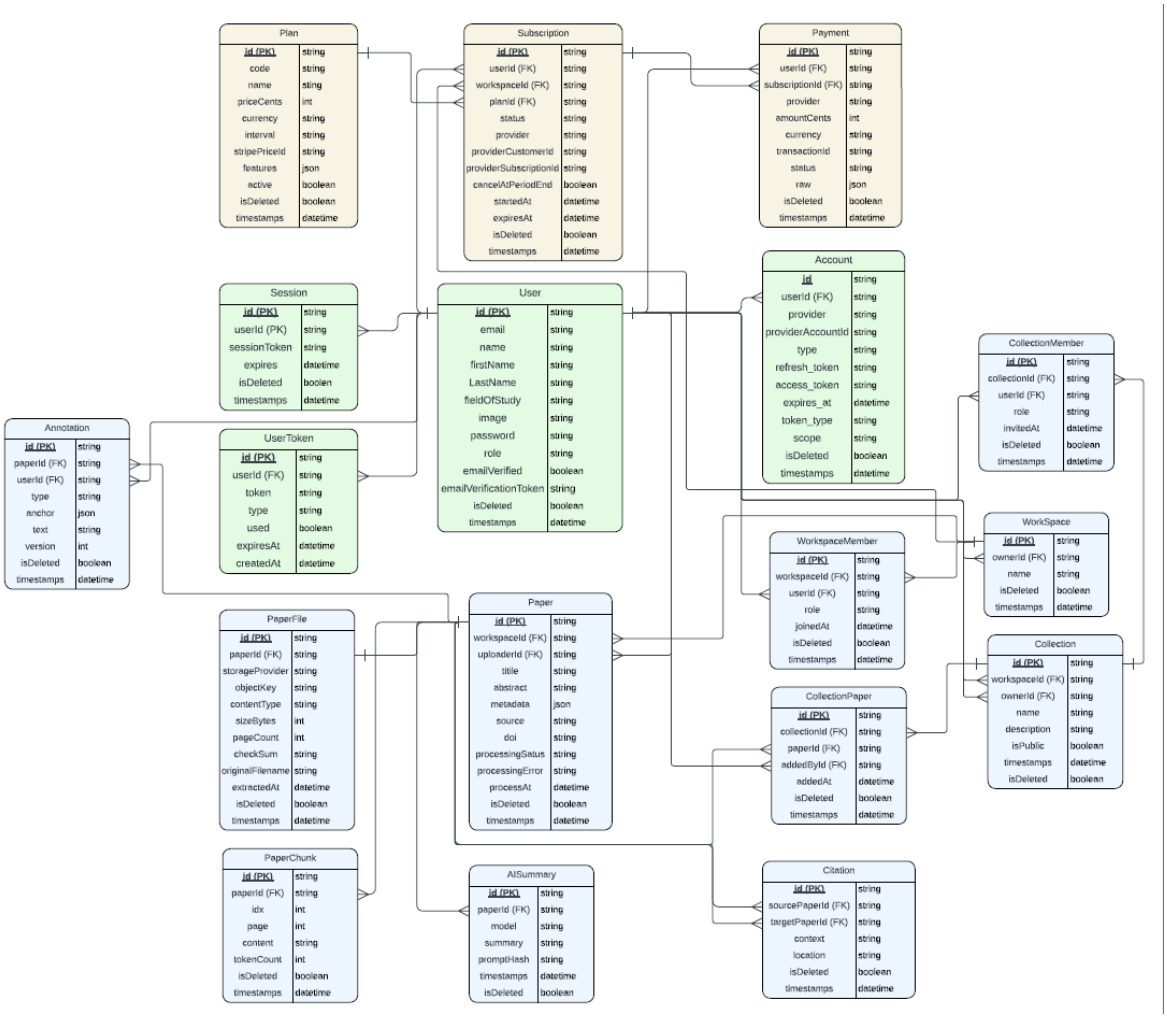


Figure 4.6: ScholarFlow Relational Schema Diagram

### 4.6.1 Core Entities (Summary)

- **Authentication**: *User*, *Account*, *Session*, and *UserToken* support local and OAuth-based authentication.
- **Collaboration**: *Workspace* and *WorkspaceMember* define collaborative boundaries and role-based membership.
- **Paper Management**: *Paper*, *PaperFile*, and *PaperChunk* manage document storage, metadata, and chunking for AI workflows.
- **Collections**: *Collection*, *CollectionPaper*, and *CollectionMember* enable structured grouping and controlled sharing.
- **AI Artifacts**: *AISummary* stores generated outputs for reproducibility and performance.

## 4.7 Sequence Diagram

The sequence diagram (Figure 4.7) captures the main interaction flow between the web client, API, database, storage, and AI provider.

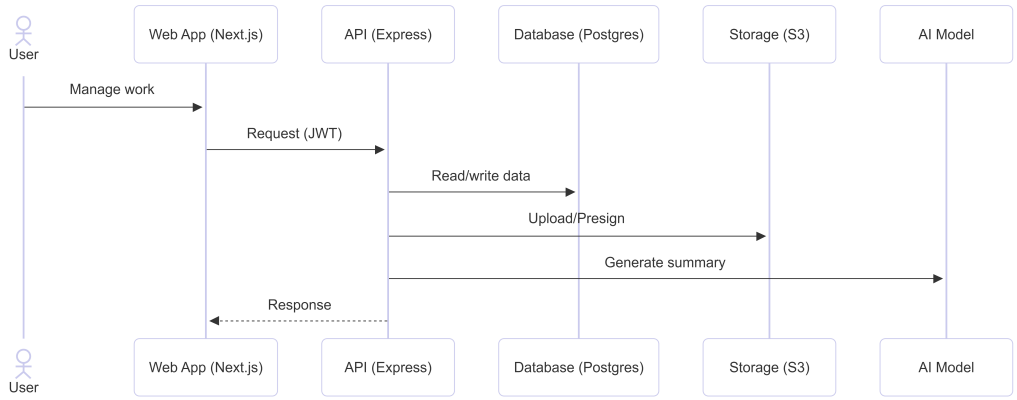


Figure 4.7: ScholarFlow Main Sequence Diagram

## 4.8 State Diagram

The state diagram (Figure 4.8) models the lifecycle of papers from upload and processing through moderation and publication.

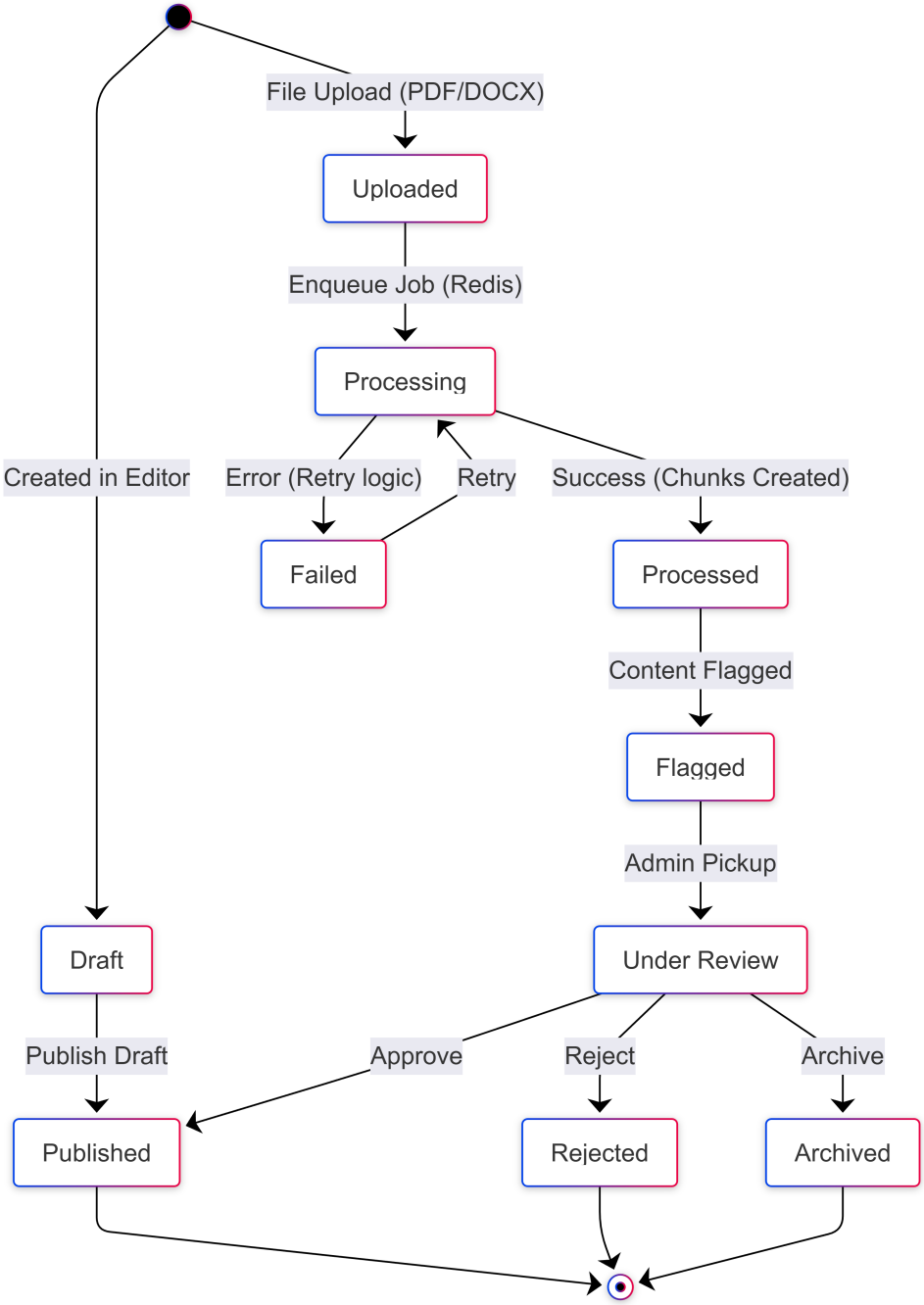


Figure 4.8: Paper Lifecycle State Diagram

## 4.9 Class Diagram

The class diagram (Figure 4.9) summarizes the core domain entities and their relationships derived from the Prisma data model.

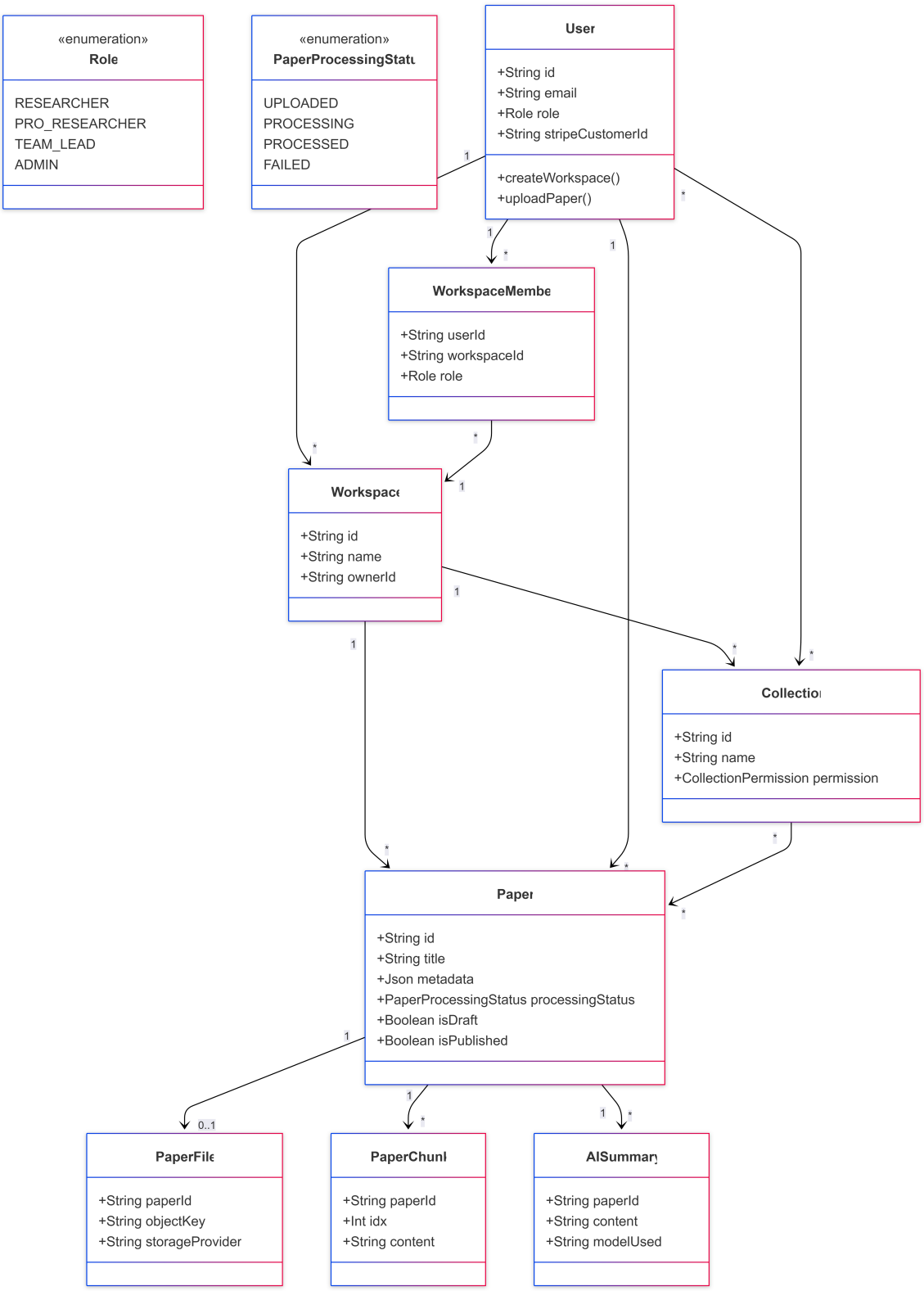


Figure 4.9: ScholarFlow Domain Class Diagram

## 4.10 CRC Cards

CRC (Class-Responsibility-Collaborator) cards capture the essential responsibilities and collaborators for primary domain classes.

### 4.10.1 User

| **Responsibility** | **Collaborator** |
|---|---|
| Authenticate and manage session | Account, Session |
| Create and own workspaces | Workspace |
| Upload and manage papers | Paper, PaperFile |
| Join workspaces as member | WorkspaceMember |
| Manage payments and subscription | Subscription, Payment |

Table 4.6: CRC Card: User

### 4.10.2 Workspace

| **Responsibility** | **Collaborator** |
|---|---|
| Container for papers and collections | Paper, Collection |
| Manage members and roles | WorkspaceMember, User |
| Handle invitations | WorkspaceInvitation |
| Track activity within scope | ActivityLog |

Table 4.7: CRC Card: Workspace

### 4.10.3 Paper

| **Responsibility** | **Collaborator** |
|---|---|
| Store metadata and content reference | PaperFile, Workspace |
| Track processing status | PaperProcessingStatus |
| Provide chunks for AI analysis | PaperChunk |
| Store AI-generated insights | AISummary, AIInsightThread |
| Maintain citation links | Citation |

Table 4.8: CRC Card: Paper

### 4.10.4 Collection

| **Responsibility** | **Collaborator** |
|---|---|
| Group papers logically | Paper, CollectionPaper |
| Manage access permissions | CollectionMember, User |
| Belong to a specific workspace | Workspace |

Table 4.9: CRC Card: Collection

# Chapter 5

# UI and Prototyping

This chapter highlights representative UI screens and prototype flows used to validate the user experience for key modules of **ScholarFlow**.

## 5.1 Authentication and Dashboard

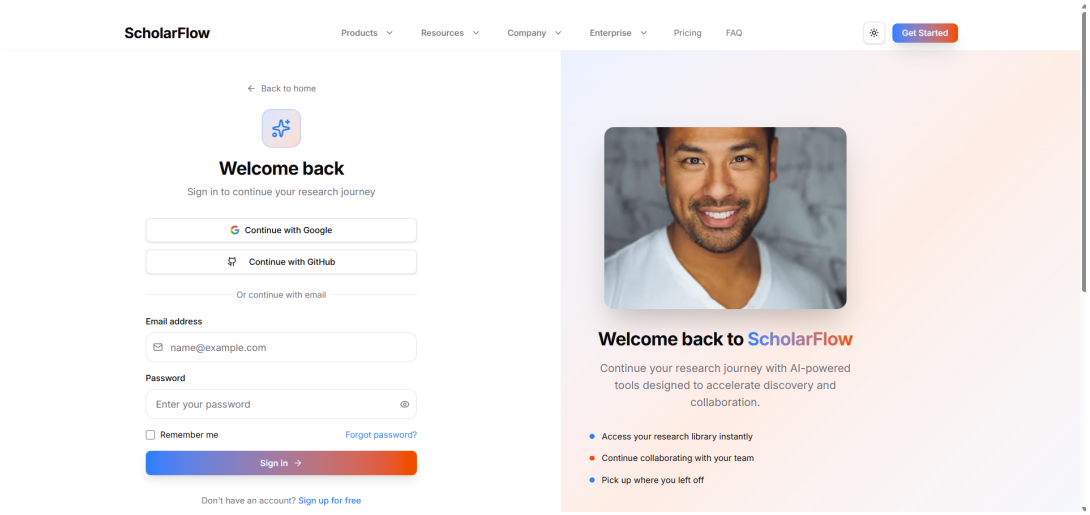


Figure 5.1: Sign In Screen

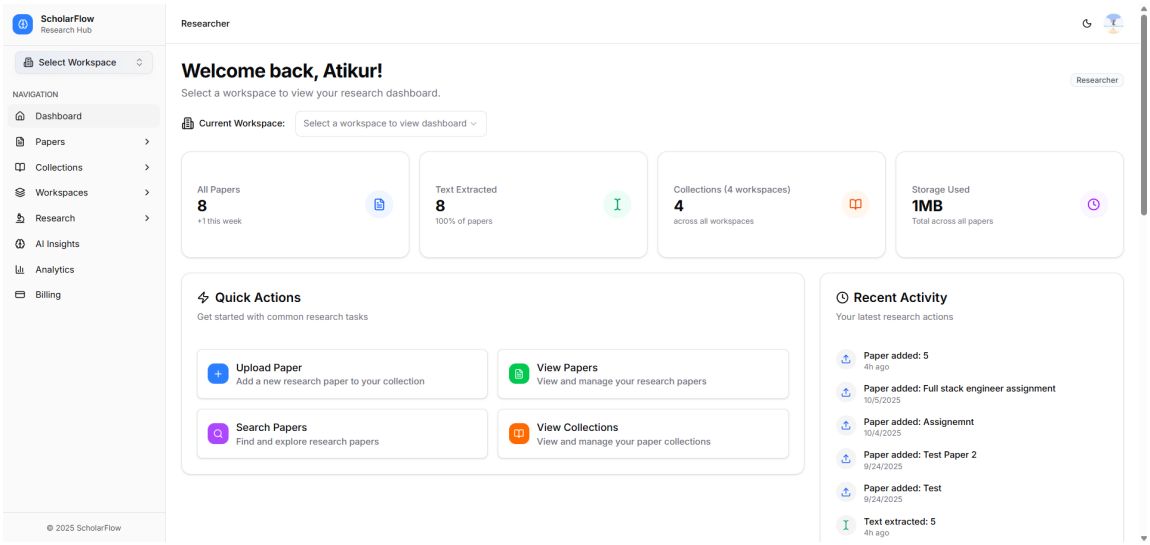


Figure 5.2: Dashboard Overview

## 5.2 Paper and Editor Workflows

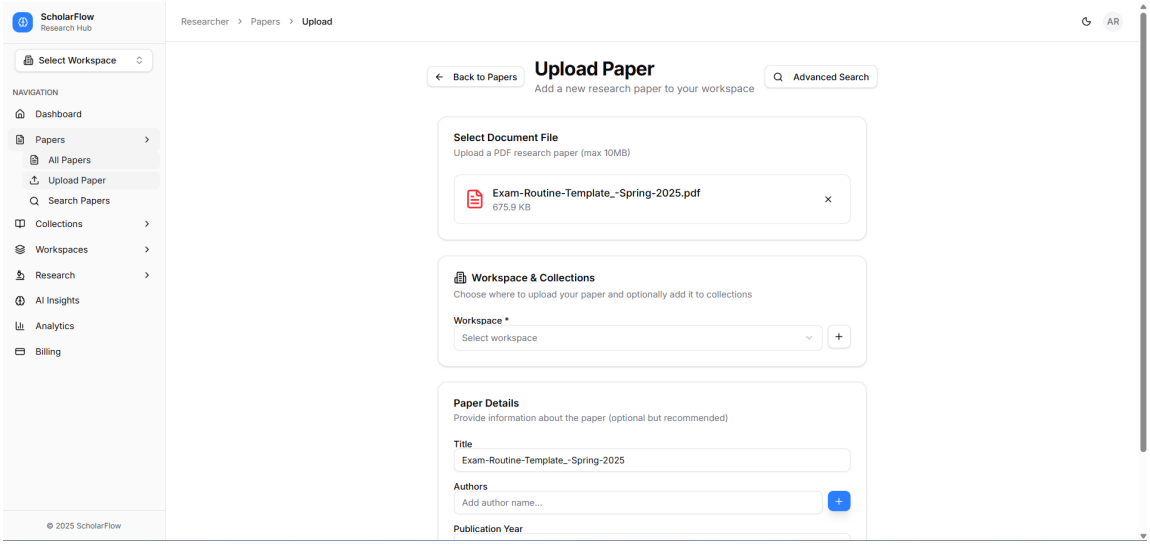


Figure 5.3: Paper Upload

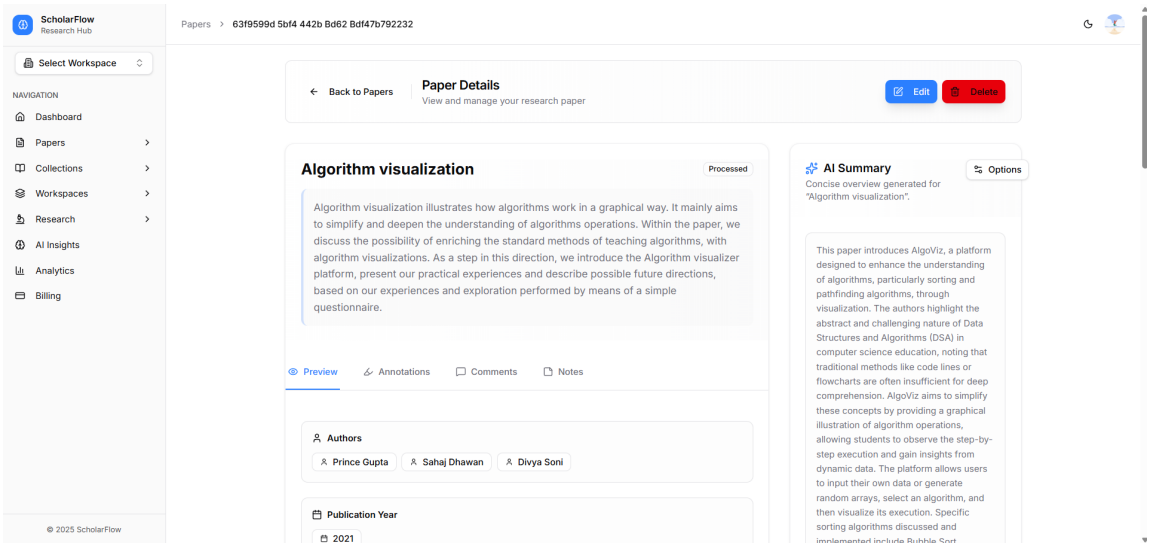


Figure 5.4: Paper Details View

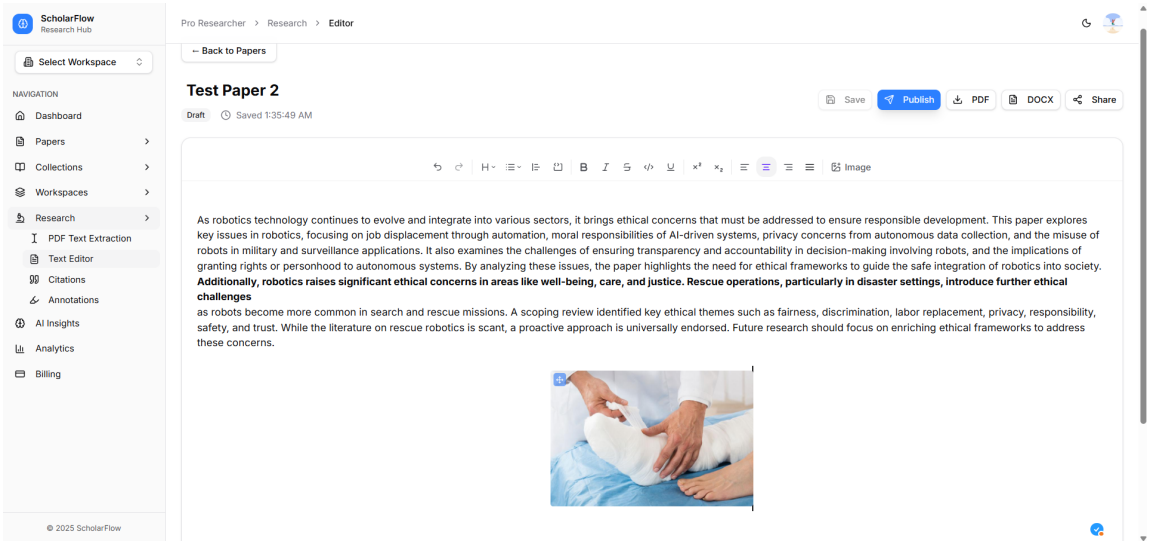


Figure 5.5: Research Editor

## 5.3 Search, AI, and Collaboration

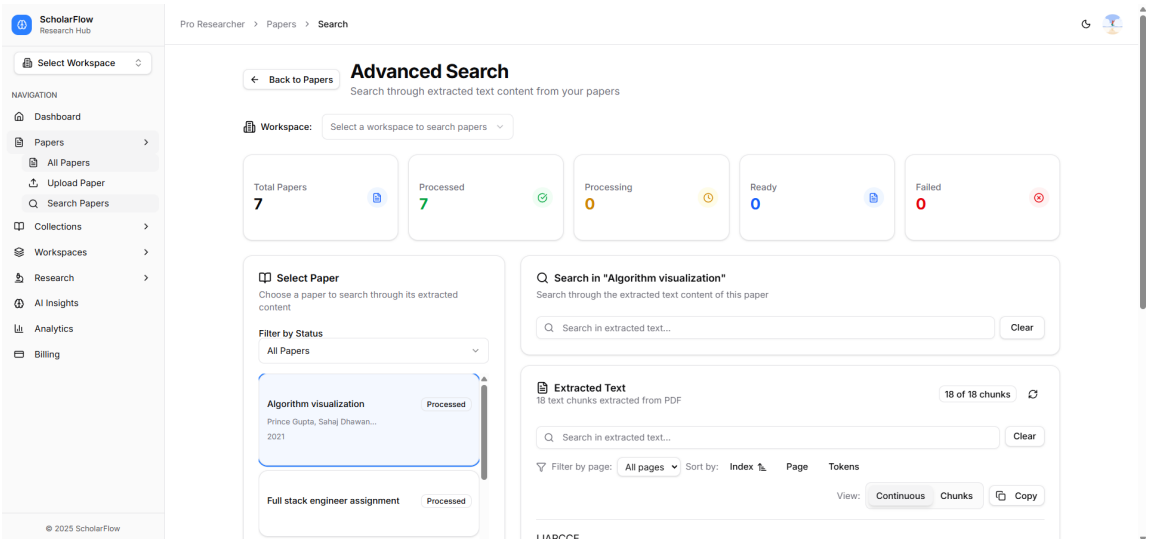


Figure 5.6: Advanced Search

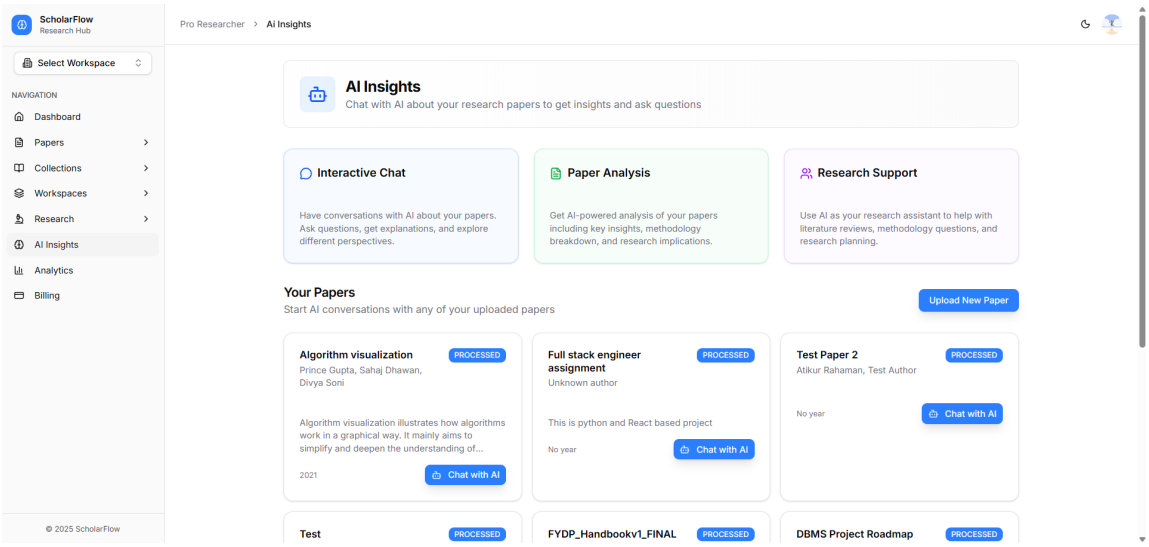


Figure 5.7: AI Insights

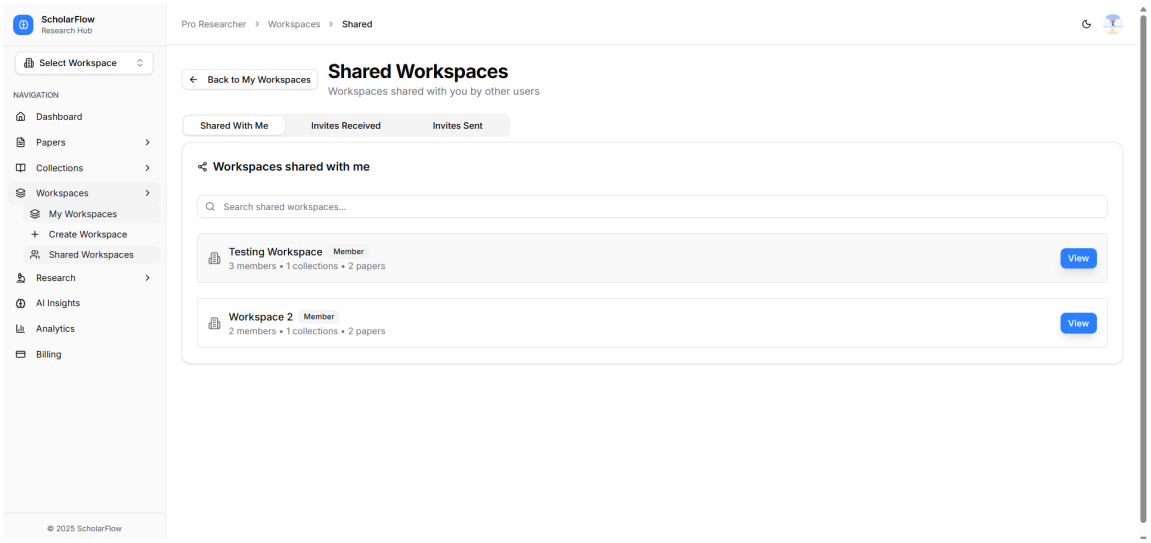


Figure 5.8: Shared Workspace

## 5.4 Collections and Billing

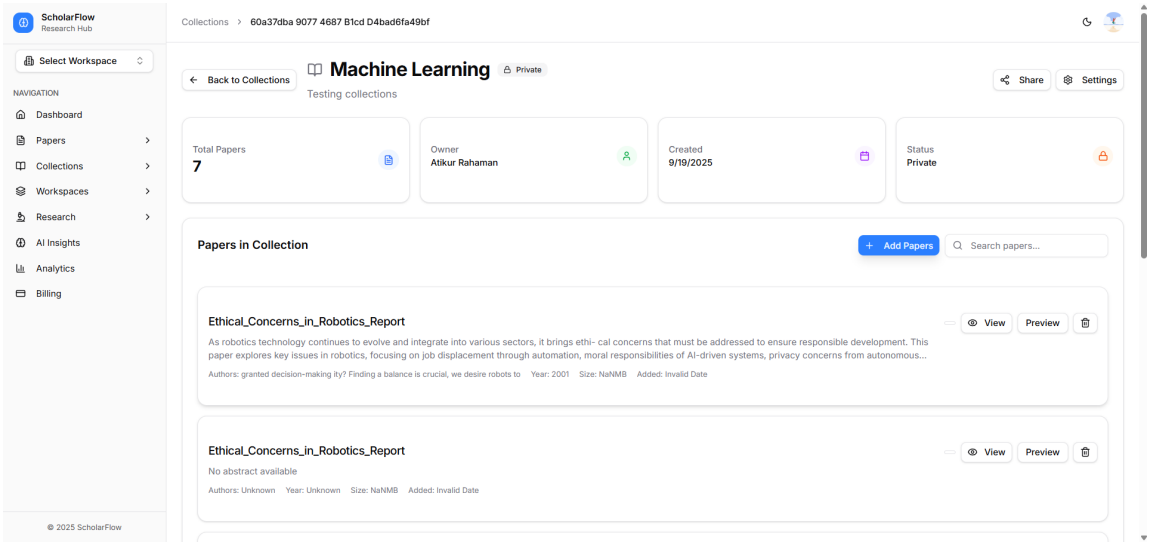


Figure 5.9: Collection Details

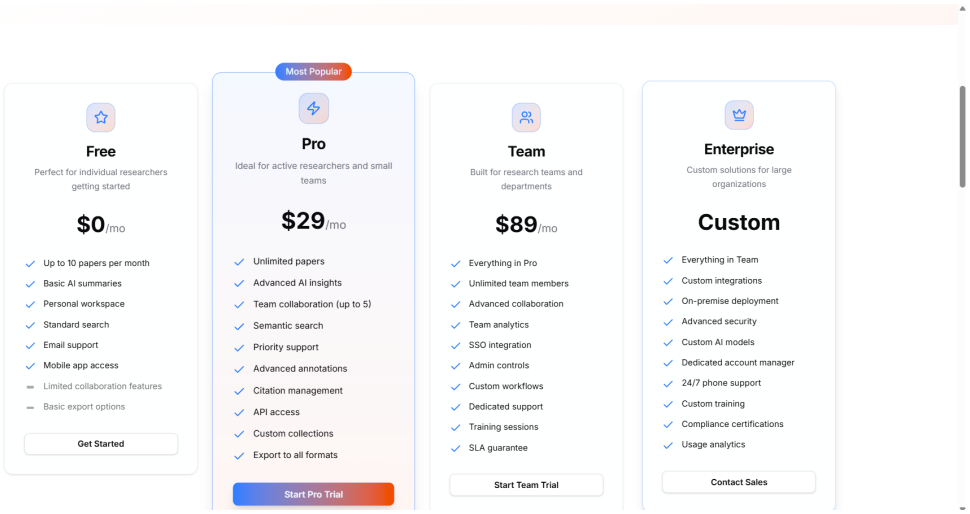


Figure 5.10: Billing Plan Selection

## 5.5 Admin and Monitoring

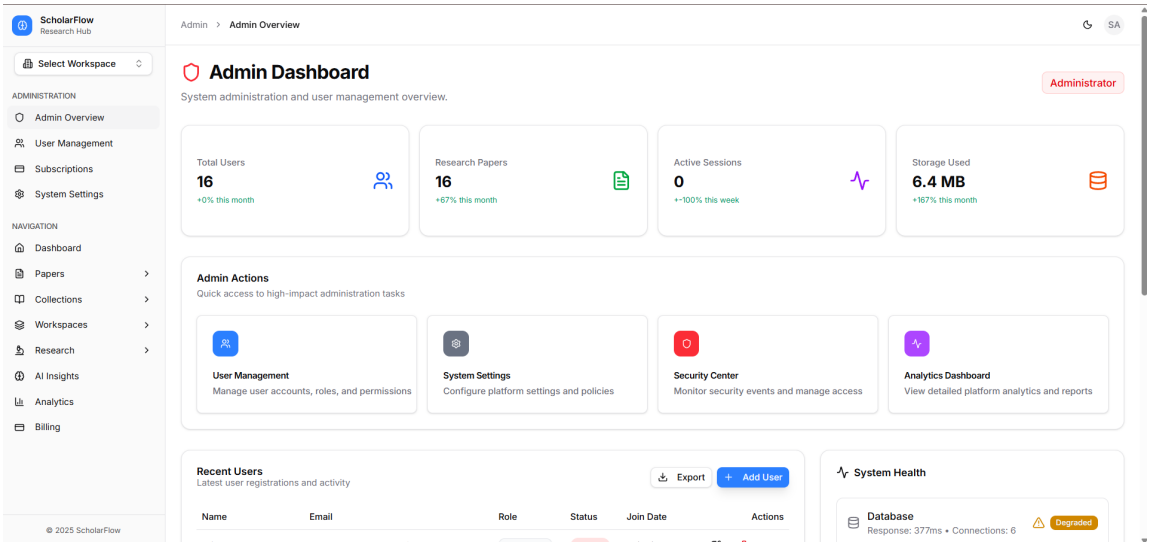


Figure 5.11: Admin Overview

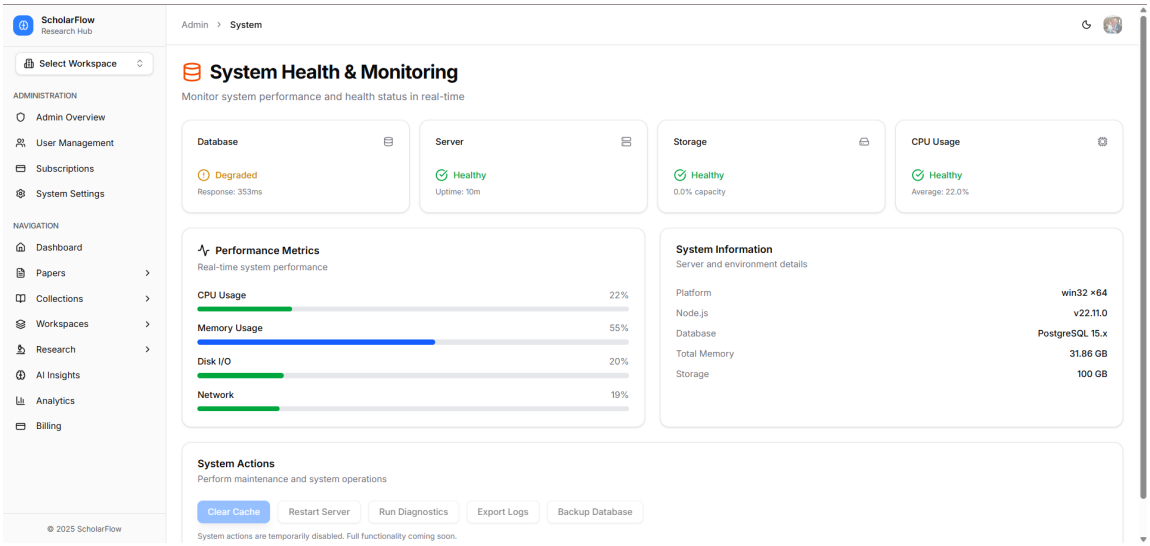


Figure 5.12: System Health Monitoring

# Chapter 6

# Conclusion

This SRS captures the requirements, system analysis, and design artifacts for **ScholarFlow**, an AI-powered research paper collaboration platform. The document formalizes functional expectations, external integrations, and core data structures while mapping them to concrete UI prototypes and architectural diagrams. The current specification provides a stable baseline for implementation, validation, and future enhancement as additional collaboration, analytics, and AI capabilities evolve.

# Chapter 7

# References


- **ScholarFlow Repository**: https://github.com/Atik203/Scholar-Flow
- **Next.js Documentation**: https://nextjs.org/docs
- **Express.js Documentation**: https://expressjs.com/
- **Prisma Documentation**: https://www.prisma.io/docs
- **PostgreSQL Documentation**: https://www.postgresql.org/docs/
- **AWS S3 Documentation**: https://docs.aws.amazon.com/s3/
- **Stripe Documentation**: https://stripe.com/docs
- **OpenAI Documentation**: https://platform.openai.com/docs
- **Google Gemini Documentation**: https://ai.google.dev/docs