AI-Powered Balance Sheet Analysis System

Submitted by: Om Agarwal
Email: agarwalom2031@gmail.com
LinkedIn: https://www.linkedin.com/in/om-agarwal-74b2aa260/
Date: 3rd August 2025

1. Introduction

This project presents a scalable and intelligent system for analyzing balance sheet documents using artificial intelligence. It enables real-time financial document interaction for analysts and executives through a chat interface powered by large language models (LLMs).

Objectives:

- Enable natural language queries on uploaded financial PDFs.
- Provide chart-based visualizations for key financial metrics.
- Enforce strict role-based access for privacy and regulatory compliance.
- Scale across companies and users with secure document segregation.

---

2. System Architecture

The system follows a modular architecture:

- **Frontend (React)**
  - Chat interface with PDF upload
  - Dropdown-toggle history panel
  - Displays answers and generated charts
- **Backend (Node.js + Express)**
  - API routes for login, chat, and file upload
  - Validates roles and filters data
  - Sends prompts to Gemini for answers
- **AI Layer (Gemini 1.5)**
  - Receives parsed PDF content
  - Answers queries and optionally returns chart-ready JSON
- **Database (MySQL)**
  - Stores users, chat history, uploaded documents
  - Relates everything via `company_id` and `user_id`

---

3. Methodology

Workflow:

1. User logs in using email/password.
2. They upload a balance sheet PDF (company-specific).
3. The PDF is parsed, indexed, and stored securely.
4. Users can ask queries like: “What are the liabilities in FY22?”
5. Based on role:
   - Analysts get data related to their company only.
   - CEOs can see all analysts' queries.

    - Admins can access everything.
6. Queries are sent to the AI model, and the response is returned to the frontend.
7. If the query involves words like “chart” or “graph”, the backend returns a JSON structure used to render a chart on the frontend.

---

4. Database Design

Tables Overview:

users

```
CREATE TABLE users (
  id INT AUTO_INCREMENT PRIMARY KEY,
  name VARCHAR(100),
  email VARCHAR(100) UNIQUE,
  password_hash VARCHAR(255),
  role ENUM('analyst', 'ceo', 'admin') NOT NULL,
  company_id INT
);
```

companies

```
CREATE TABLE companies (
  id INT AUTO_INCREMENT PRIMARY KEY,
  name VARCHAR(100) NOT NULL,
  parent_company_id INT DEFAULT NULL,
  FOREIGN KEY (parent_company_id) REFERENCES companies(id)
);
```

balance_sheets

```
CREATE TABLE balance_sheets (
  id INT PRIMARY KEY AUTO_INCREMENT,
  company_id INT,
```

```
  year INT,
  revenue DECIMAL(15,2),
  profit DECIMAL(15,2),
  assets DECIMAL(15,2),
  liabilities DECIMAL(15,2),
  growth DECIMAL(5,2),
  created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
  FOREIGN KEY (company_id) REFERENCES companies(id)
);
```

chat_history

```
CREATE TABLE chat_history (
  id INT AUTO_INCREMENT PRIMARY KEY,
  user_id INT,
  company_id INT,
  question TEXT,
  answer TEXT,
  response_type ENUM('text', 'chart', 'image') DEFAULT 'text',
  created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
  FOREIGN KEY (user_id) REFERENCES users(id),
  FOREIGN KEY (company_id) REFERENCES companies(id)
);
```

This schema enforces:

- Role-based access
- Company-specific data segregation
- Chat traceability for future audit or insights

---

5. Key Findings

- Users can interact with complex financial PDFs via simple queries.
- Role-based filtering works reliably for data privacy.
- Gemini can generate concise, accurate answers.
- Chart support (via keywords) makes visualization intuitive.
- PDF upload and parsing support a flexible document pipeline.

---

6. Limitations

- Charts rely on simple JSON instructions; complex visuals aren't supported yet.
- PDF parsing accuracy drops if documents are unstructured or scanned.
- Load balancing not yet implemented for horizontal scaling.
- No support yet for multi-language queries.

---

7. Future Enhancements

- Add downloadable PDF summaries per query.
- Include audit logs for compliance purposes.
- Expand to analyze cash flow and P&L statements.
- Integrate admin dashboard with analytics.
- Deploy serverless architecture for scalability.

---

8. Conclusion

The Balance Sheet Analysis System delivers a practical, AI-driven solution to financial document analysis. By combining NLP, visualizations, and granular access control, it sets the foundation for smart financial reporting tools tailored to enterprise use cases.