Knowledge Management and Assessment System Based on AI

Authors: Zihong Luo (ZL), Qiaoyang Bi (QB), Tianyu Zhu (TZ),
Sijin Wu (SW), Pengyu Lu (PL), Dawei Wang (DW)

I. Project Description

This project develops an advanced PHP-MySQL-based knowledge management and self-assessment platform. The system supports multi-format file uploads—including PDF, DOCX, and TXT—and integrates DeepSeek AI to perform automatic summarization, extract critical knowledge points, and generate assessment questions. The platform is designed to store historical data, organize knowledge effectively, facilitate file downloads, and track user progress for in-depth analysis. Key functionalities include robust file management, AI-driven content processing, structured knowledge organization, comprehensive user access control, and stringent data security measures. All these features are presented through an intuitive and user-friendly interface, making it an ideal tool for enhancing learning efficiency and assessment accuracy. The system is branded as *ai deepchat*.

II. Aims and Objectives

Aim:

To develop an intelligent platform that seamlessly integrates document management, automated knowledge extraction, and assessment generation. The platform aims to enhance the quality and efficiency of learning and training processes by automatically summarizing, extracting, and organizing key knowledge points from a variety of documents, and subsequently generating relevant assessment questions.

Objectives:

a. **Efficient Document Management**:

- Develop a robust mechanism for file uploads and downloads that supports multiple file formats.
- Implement version control to track and manage changes in documents over time.
- Ensure secure storage and quick retrieval of documents.

b. **Intelligent Knowledge Point Extraction**:

- Leverage DeepSeek AI to automatically analyze uploaded documents.
- Identify and extract critical knowledge points, reducing the need for manual intervention.
- Enhance accuracy through iterative learning and refinement of AI algorithms.

c. **Automated Assessment Question Bank Generation**:

- Generate diverse assessment questions—including multiple-choice, true/false, and fill-in-the-blank formats—from the extracted knowledge points.
- Create a dynamic, reusable question bank that can adapt to different learning levels and topics.

d. **User-Friendly Frontend**:

- Design an intuitive web interface using PHP, HTML, CSS, and JavaScript.
- Ensure ease-of-use for tasks such as file management, knowledge point editing, and assessment administration.

e. **History Records and Version Tracking**:

- Maintain a comprehensive MySQL database to log user edits, document versions, and summary records.
- Enable users and administrators to compare different versions for audit and quality control.

f. **Role-Based Permissions and Data Security**:

- Implement strict role-based access control to manage different user types (e.g., administrators, regular users, and visitors).
- Enforce robust data security measures to protect sensitive information.

g. **Scalability and Maintainability**:

- Build the platform using a mature and scalable technology stack.
- Lay the groundwork for future enhancements, such as advanced recommendation systems and upgraded AI algorithms.

h. **Multidimensional Feedback and Analysis**:

- Provide comprehensive analysis of assessment results.
- Offer actionable learning feedback to users or organizations for continuous improvement.

III. Key Literature and Background Reading

Key Sources Include:

- API documentation for DeepSeek AI integration to understand technical requirements and limitations.
- Recent research papers on automated question generation to inform the design of the assessment module.

- Comprehensive literature on knowledge management systems and learning analytics to benchmark best practices.
- Studies focused on role-based access control and data security to ensure that the platform adheres to modern security standards.

Background Reading:

A: ChatGPT is recognized as an effective text summarization tool, providing practical solutions for extracting key information in a report-style format (Sarraf and Abbaspour, 2023).

B: Research on Claude AI indicates that AI can significantly alleviate the workload of librarians by automating subject analysis (Martins, 2024).

C: Technical Evaluation Report examines the complementary use of Notion and Notion AI in enhancing second language writing processes (Osawa, 2023).

D: AI Summary Report presents evidence that AI-generated summaries can improve factual knowledge retention, independent of user awareness of AI involvement (Karell et al., 2025).

E: Research Paper underscores the importance of automatic text summarization in managing and extracting information from large data sets (Day and Chen, 2018).

IV. Development and Implementation Summary

The system architecture is modular and comprises the following key components:

- **Back-end**: Built on PHP and MySQL, the back-end manages secure data storage, server-side logic, and integration with AI services.
- **Front-end**: Developed using HTML, CSS, and JavaScript, the front-end ensures an intuitive, responsive user interface that facilitates smooth navigation and interaction.
- **AI Integration**: DeepSeek AI is integrated to analyze document content, extract salient information, and generate corresponding assessment questions. Continuous updates and model training ensure the system adapts to new data and requirements.
- **Security Module**: This includes role-based authentication, data encryption, and compliance with modern data protection standards to safeguard sensitive user information.

Workflow:

1) A user uploads a document through the secure interface.
2) The AI engine processes the document to extract key knowledge points.
3) Based on these knowledge points, the system automatically generates a set of assessment questions.
4) Users can review, organize, and, if necessary, edit the extracted knowledge points.
5) Finally, users undertake assessments, and their performance data is recorded for further analysis and feedback.

V. Data Sources

DeepChat AI utilizes a diverse range of data sources to ensure robust and accurate knowledge extraction:

- **User-Uploaded Documents**: These serve as the primary input and support various file formats such as PDF, TXT, and DOCX.
- **Public Domain Datasets**: Academic papers, technical manuals, and other publicly available resources are integrated to enhance the AI model's accuracy.
- **AI-Processed Knowledge Points**: Extracted knowledge is systematically stored for efficient organization, retrieval, and future reference.
- **Assessment Data**: Data from generated test questions and user responses is collected to perform learning analytics and further refine the AI algorithms.

All data is processed in strict accordance with established security and privacy policies to ensure confidentiality and compliance.

VI. Testing and Evaluation

To ensure DeepChat AI operates effectively, multiple testing strategies will be implemented:

- **Unit Testing**: Verifies the accuracy and reliability of backend functions and AI-driven knowledge extraction and question generation.
- **UI/UX Testing**: TestingGathers user feedback to improve interface design, accessibility, and overall user experience.
- **Performance Evaluation**: Assesses the relevance and accuracy of AI-generated questions while monitoring system response times and efficiency.
- **Security Testing**: Evaluates data encryption, access control, and potential vulnerabilities to ensure robust data protection and user privacy.

These tests collectively enhance system stability, usability, and security.

VII. BCS Project Criteria

- Application of practical and analytical skills in AI and web development.
- Innovative approach to automated knowledge assessment.
- Integration of AI techniques with learning analytics.
- Addressing a real-world need for efficient knowledge management.
- Effective self-management of project timelines and milestones.
- Critical evaluation of the system's effectiveness and potential improvements.

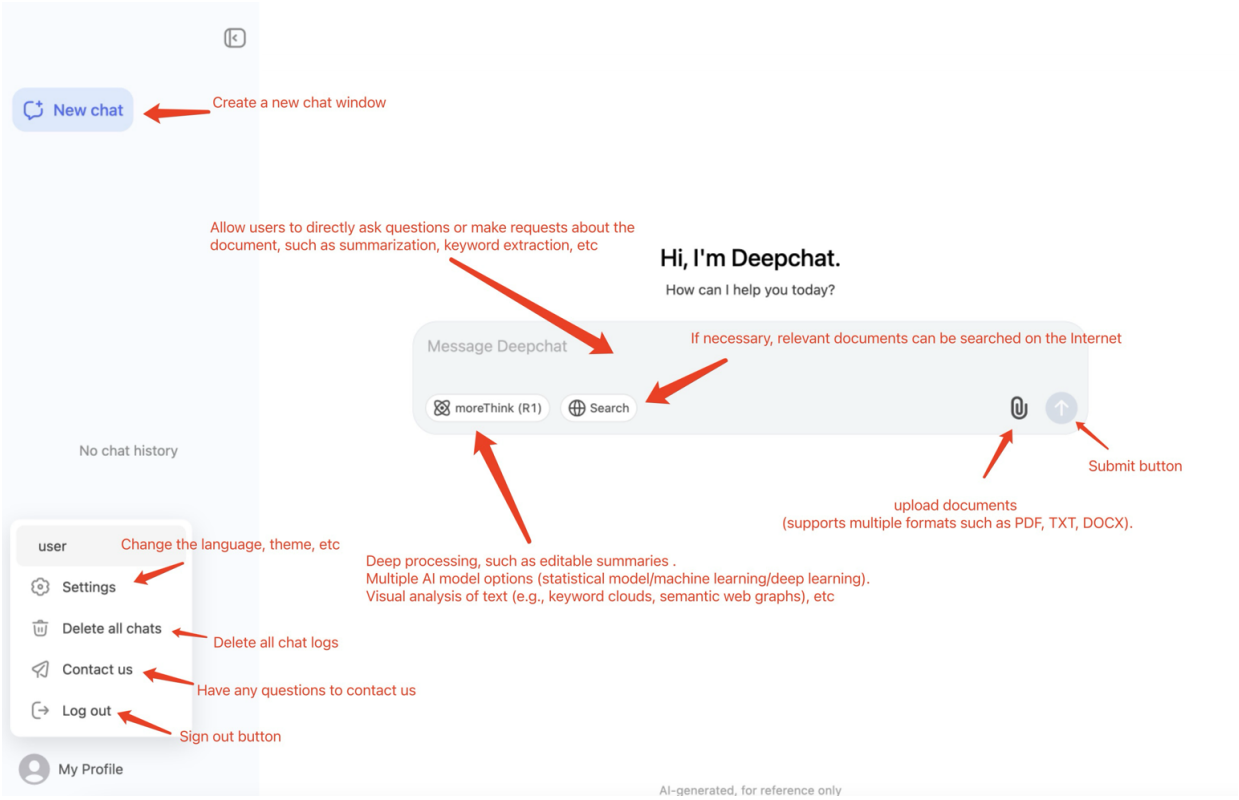
VIII. UI/UX Mockup

User Interface (UI):

- **Input Field**: Allows users to paste text or upload documents (supports multiple formats such as PDF, TXT, DOCX).
- **API Access Options**: Provides developers with API access options.
- **Summary Generation Options**:
  - Summary length selection (short, medium, long).
  - Summary style selection (formal, academic, casual, technical).
  - Language selection (multi-language support).
- **Submit Button**: Includes progress indicators or loading animations to show processing status.
- **Results Display**: Presents results based on the generated summaries and questions.

User Experience (UX):

- Automatic recognition of input formats to handle various text types.
- Real-time preview of the summary generation process to reduce user anxiety.
- Adjustable summary weighting, allowing users to fine-tune key information.
- Custom AI training options where user feedback is used to refine AI-generated summaries.
- Comparison of word frequencies between the original text and the generated summary.
- Color-coded highlights to emphasize key information for quick comprehension.
- Summary quality score charts to help users evaluate the effectiveness of the AI-generated summary.



Gambar 1. UI/UX Mockup

IX. Risk and Contingency Planning

A. Technical Risks

- **System Downtime**: Utilize cloud backups, server redundancy, and load balancing.
- **Database Corruption**: Implement automated backups, data integrity checks, and recovery mechanisms.
- **Performance Bottlenecks**: Address through code optimization, caching strategies, and load testing.
- **AI Generation Errors**: Provide manual correction options, integrate user feedback, and employ backup AI models.

B. Security Risks

- **Data Breach**: Prevent via end-to-end encryption, access control, and regular security audits.
- **Cyber Attacks**: Employ firewalls, regular security patch updates, and intrusion detection systems.
- **Insider Threats**: Use role-based access control along with activity logging and tracking.

C. Development and Maintenance Risks

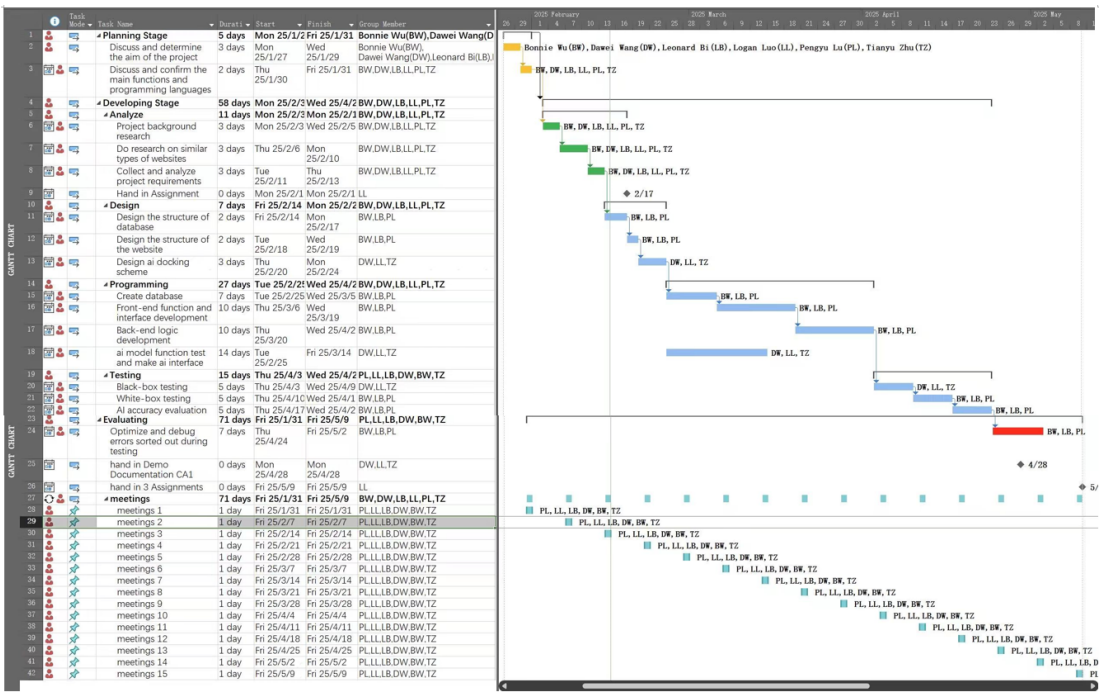
- **Scope Creep**: Mitigate through clear requirements, agile development, and milestone tracking.
- **Third-Party API Failures**: Prepare backup AI solutions and automatic API fallback mechanisms.
- **Software Bugs**: Address with automated testing, continuous integration (CI/CD), and code reviews.

D. User-Related Risks

- **Low User Engagement**: Enhance through an intuitive UI/UX, interactive learning, and personalized recommendations.
- **Unauthorized Content Uploads**: Mitigate with AI-based content filtering, manual moderation, and policy enforcement.
- **User Retention Issues**: Improve by tracking learning progress, implementing reward systems, and integrating user feedback.

E. Operational Risks

- **Scalability Limitations**: Address using cloud scaling, a microservices architecture, and load balancing.
- **Regulatory Compliance Issues**: Ensure compliance with GDPR and data protection regulations, and provide robust user privacy controls.



Gambar 2. Gantt Chart

X. Reference List

1) Day, M.-Y. and Chen, C.-Y. (2018). *Artificial Intelligence for Automatic Text Summarization*. In: 2018 IEEE International Conference on Information Reuse and Integration (IRI). IEEE. doi:10.1109/iri.2018.00076.
2) Karell, D., Shu, M., Okura, K. and Davidson, T. (2025). *Artificial Intelligence Summaries of Historical Events Improve Knowledge Compared to Human-Written Summaries*. Available at: https://osf.io/preprints/socarxiv/3gsqw_v1 [Accessed 14 Feb. 2025].
3) Martins, S. (2024). *Artificial Intelligence-Assisted Classification of Library Resources: The Case of Claude AI*. Library Philosophy and Practice, E-journal(8159), 8159. Available at: ResearchGate Publication.
4) Osawa, K. (2023). *Integrating Automated Written Corrective Feedback into E-Portfolios for Second Language Writing: Notion and Notion AI*. RELC Journal, 55(3). doi:10.1177/00336882231198913.
5) Sarraf, A. and Abbaspour, A. (2023). *ChatGPT Application in Summarizing an Evolution of Deep Learning Techniques in Imaging: A Qualitative Study*. arXiv. Available at: https://arxiv.org/abs/2312.03723 [Accessed 14 Feb. 2025].