IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 6, Ser. 4 (Nov. – Dec. 2025), PP 12-17
www.iosrjournals.org

AI-Based Automated Flashcard Generator

Dr Hema Jagadish[1], Manya D M[2], Medha Bellur[3], Prerana G S[4], Jagruthi B J[5]

(Department Of Ise/ Bangalore Institute Of Technology, India)
(Department Of Ise/ Bangalore Institute Of Technology, India)

Abstract:
The AI Flash Card Generator is an innovative application designed to automate the creation of digital flash cards using artificial intelligence. It leverages natural language processing (NLP) techniques to extract key concepts, definitions, and questions from educational content such as textbooks, notes, or articles. The system aims to reduce manual effort in preparing study material and enhance learning efficiency for students and professionals. By generating question–answer pairs, it supports active recall and spaced repetition, two proven strategies for memory retention. The generator also allows customization, enabling users to select difficulty levels, subjects, or formats. With an intuitive interface, learners can easily review, edit, or share flash cards across devices. The integration of AI ensures context-aware, accurate, and concise content generation. The project enhances self-learning, exam preparation, and digital education tools. It demonstrates the role of AI in personalized learning solutions. Ultimately, the AI Flash Card Generator serves as a smart assistant for effective knowledge reinforcement.

Key Words: *Artificial Intelligence; Flashcard Generation; PDF Text Extraction; Optical Character Recognition (OCR); BLIP Image Captioning; Educational Technology; Interactive Learning; Automated Study System; Learning Personalization; Question-Answer Generation*

Date of Submission: 09-12-2025 Date of Acceptance: 19-12-2025

I. Introduction

The AI Flashcard Generator is an intelligent tool designed to automatically create digital flashcards from raw text, documents, or study materials. Flashcards are widely used in education for active recall and spaced repetition, two proven methods for effective learning and memory retention. Traditionally, students or teachers manually prepare flashcards, which is a time-consuming and repetitive process. With the help of Artificial Intelligence (AI) and Natural Language Processing (NLP), tasks can be automated. The AI Flashcard Generator analyzes input text, identifies important concepts, definitions, and key information, and then converts them into a structured question–answer format suitable for flashcards. In the proposed project an AI-powered flashcard generator is built using Python. The system will utilize libraries like NLTK, SpaCy, and Hugging Face Transformers for natural language processing, along with PyPDF2 or pdfplumber for extracting text from documents. The generated flashcards can be stored in structured formats (JSON, CSV, or database) and later integrated into learning platforms or mobile apps.

II. Literature Survey

Referential literature refers to scholarly or academic works that are cited or referenced within other texts to support arguments, provide background, or acknowledge prior research. It serves as a foundation for new studies, ensuring credibility, continuity, and context in academic writing. Referential literature is essential in research as it connects current work with established knowledge. In the context of an AI flashcard generator, referential literature plays a crucial role in ensuring the accuracy, credibility, and educational value of the generated content.

1) Pikle, Shlok, et al. "FlashMe: Automatic Flashcard Generation." 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, 2023, vol. 2023, pp. 1–6.

The study addresses the problem of inefficient study techniques commonly used by students by proposing the use of advanced Natural Language Processing (NLP) methods. The researchers designed a multi-stage transformer architecture that can automatically generate flashcards from text containing dense, technically rich information, such as textbook material.

While the approach shows promise, it comes with certain limitations. Transformer-based NLP models generally require significant computational resources, which may restrict their accessibility in low-resource environments. Moreover, the quality of the generated flashcards is highly dependent on the accuracy and contextual understanding of the model, meaning some flashcards may oversimplify or omit critical details.

2) Thabet, et al. "Q&A Generation for Flashcards Within a Transformer-Based Framework." Communications in Computer and Information Science, vol. 1812, Springer, 2023, pp. 789–806.

The study proposes an innovative system that dynamically generates flashcards—comprising supporting paragraphs, questions, and answers—by integrating transformer-based language models with a recommender system tailored for learning programming concepts. The foundation lies in an extension of a previous Intelligent Serious Games (ISG) model, which originally combined Deep Knowledge Tracing (DKT) with a Transformer-based recommender to anticipate learner outcomes and suggest flashcards to aid in progressing through game missions.

Despite its promising capabilities, the proposed transformer-based flashcard generation framework presents several limitations. First, fine-tuning and deploying large transformer models like GPT-2, BART, and T5 typically require considerable computational resources, which may hinder adaptation in resource-constrained educational settings. Secondly, while standard coherence and semantic metrics assess textual quality, they do not necessarily guarantee pedagogical value—questions or answers may still lack depth or alignment with learning objectives.

3) Cheng, Yuang, et al. "WikiFlash: Generating Flashcards from Wikipedia Articles." Lecture Notes in Computer Science, vol. 12908, Springer, 2021, pp. 138–149.

The researchers introduced WikiFlash, a system designed to automatically create flashcards directly from Wikipedia articles, aiming to support self-learners by reducing the tedious effort of manual flashcard creation. They implemented a multi-stage pipeline leveraging state-of-the art NLP and transformer-based models. These models handle tasks including summarization, question generation, and answer extraction to form structured question–answer pairs

Despite its promising approach, the WikiFlash system has several limitations. First, while factoid-style flashcards are straightforward to generate, more complex, inference-based questions might still pose a challenge due to limitations in model reasoning capabilities—such difficulty is acknowledged in related research on neural question generation. Additionally, transformer models typically demand substantial computational resources for both training and inference, which might hinder deployment in low-resource environments.

4) Bachiri, et al. "Optimizing Learning Outcomes and Retention in MOOCs with AI-Generated Flashcards." Lecture Notes in Educational Technology, vol. 358, Smart Learning for a Sustainable Society, Springer Nature, 2023, pp. 242–247.

The study presents a Flashcard Bot that harnesses generative AI within the Massive Open Online Courses (MOOCs) context to improve student learning and retention. The system uses the pretrained T5 model, a widely used transformer-based language model, to convert extracted course content from a business intelligence MOOC hosted on Moodle into flashcards. The method aims to leverage cognitive learning principles—specifically the testing effect and spaced repetition—by generating flashcards that help learners effectively review and retain information over time.

Despite the encouraging outcomes, the approach has several notable limitations. First, deploying and fine-tuning a model like T5 typically requires significant computational resources, which may hinder scalability or adoption in resource-constrained learning environments. Although the AI-generated flashcards are evaluated favorably by students in terms of perceived quality and fluency, the study doesn't appear to assess actual learning outcomes

5) Ghosh. "Interactive Augmented Reality Application Using Animal Flashcards for Education of Children." Taylor & Francis, 2025.

In the study, A. Ghosh introduces an innovative marker-based Augmented Reality (AR) application that brings animal flashcards to life in a playful learning environment. The app features a "Guess the Animal" game using voice recognition: children verbally identify the animal displayed through AR, and the system confirms whether the answer is correct using a speech module. his interactive setup aims to engage young learners more effectively by combining visual representation, audio feedback, and game mechanics.

Despite its engaging design, the approach carries several potential limitations. Firstly, marker based AR systems often depend heavily on precise lighting and positioning—factors that can be unpredictable in typical learning environments, which may lead to unreliable performance or frustration among young users. Secondly, speech recognition technologies, especially with children's voices, can struggle with pronunciation variability and ambient noise, potentially decreasing usability.

III. Objectives

- To design and develop a system that can automatically generate flashcards from raw text or documents.
- To apply Natural Language Processing (NLP) techniques for extracting key concepts and generating meaningful question–answer pairs.
- To store the generated flashcards in reusable and structured formats such as JSON, CSV, or text files.
- To provide students and educators with a user-friendly tool that enhances active recall and efficient revision.
- To explore the scope for future extension into a web or mobile-based application with advanced features.

IV. Problem Statement

In today's education system, students must manage and retain large volumes of information from various sources, making traditional methods like notetaking and memorization less effective and time-consuming. Flashcards have proven to be a valuable learning tool by promoting active recall and spaced repetition, yet manually creating these flashcards can be tedious, especially for extensive material. Although digital flashcard apps such as Quizlet and Anki exist, they still depend on users to input data manually, which can limit their usefulness. With advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP), it is now possible to automate flashcard creation. AI can analyze text, identify key concepts, and generate question–answer pairs, significantly reducing the time and effort students spend preparing study materials. It not only streamlines the learning process but also allows for more personalized study aids tailored to individual learning needs. Automating flashcard generation can improve knowledge retention and engagement by consistently delivering well-structured, relevant content. The project aims to leverage these technologies by developing an AI-powered flashcard generator using Python, helping students focus more on learning and less on content preparation.

V. System Architecture

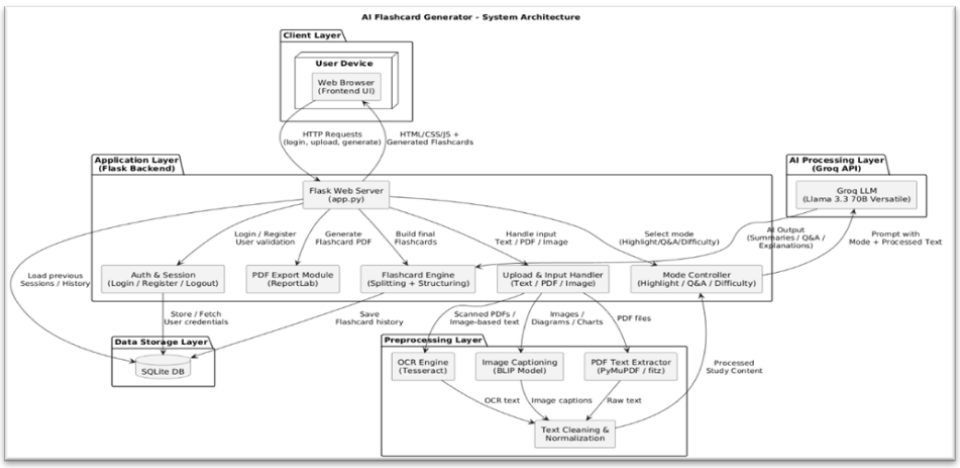

Figure 1. System Architecture of AI Flashcard Generator

The system architecture illustrates a complete workflow for an AI-powered flashcard generator, connecting client interaction, backend processing, AI inference, and storage operations. A web browser on a user device sends requests to a Flask server, which manages authentication, PDF handling, flashcard generation, and export functions. Incoming study material—whether text, PDF, or images—passes through preprocessing modules such as OCR, image captioning, text extraction, and normalization before reaching the core flashcard engine. A mode controller guides the system to produce summaries, Q&A pairs, explanations, or difficulty-based content. Processed input is forwarded to a Groq-based LLM for refined outputs, which return to the server for structuring and presentation on the frontend. An SQLite database stores login information, session history, and previously generated flashcards, enabling smooth retrieval and continuity across sessions.

VI. Methodology

The methodology to be followed in developing the AI Flashcard Generator is as follows:

- **Requirement Analysis–**Identify the objectives, software, and hardware requirements needed for implementation. [2]
- **Data Collection–**Gather input study material in the form of text, PDFs, or notes.
- **Preprocessing–**Perform text cleaning, tokenization, and stop word removal to prepare the data for analysis.
- **Concept Extraction–**Apply NLP techniques to extract key terms and definitions from the processed text.
- **Flashcard Generation-**Convert extracted concepts into Key conceptual points and customized question–answer pairs using AI/NLP models.[4]
- **Storage and Organization**–Save the flashcards in reusable formats such as JSON or CSV for easy retrieval.
- **User Interaction–**Provide the flashcards in a structured and user-friendly manner for study and revision.
- **Testing and Evaluation–**Evaluate the accuracy, relevance, and clarity of generated flashcards, and refine the system based on feedback.

VII. Results

The system architecture illustrates a complete workflow for an AI-powered flashcard generator, connecting client interaction, backend processing, AI inference, and storage operations. A web browser on a user device sends requests to a Flask server, which manages authentication, PDF handling, flashcard generation, and export functions. Incoming study material—whether text, PDF, or images—passes through preprocessing modules such as OCR, image captioning, text extraction, and normalization before reaching the core flashcard engine. A mode controller guides the system to produce summaries, Q&A pairs, explanations, or difficulty-based content. Processed input is forwarded to a Groq-based LLM for refined outputs, which return to the server for structuring and presentation on the frontend. An SQLite database stores login information, session history, and previously generated flashcards, enabling smooth retrieval and continuity across sessions.

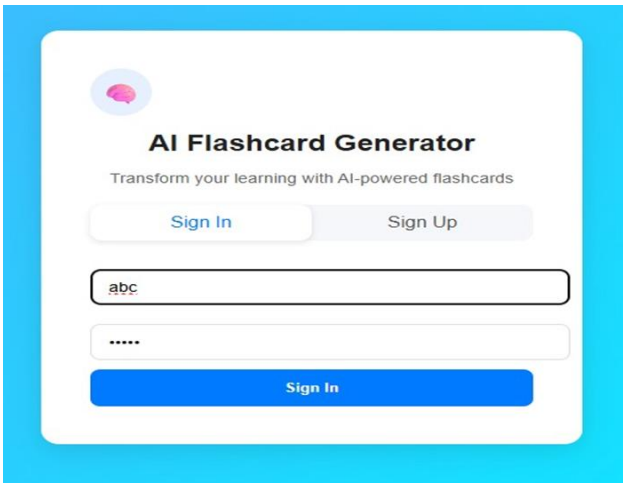

Figure 1: Login Page

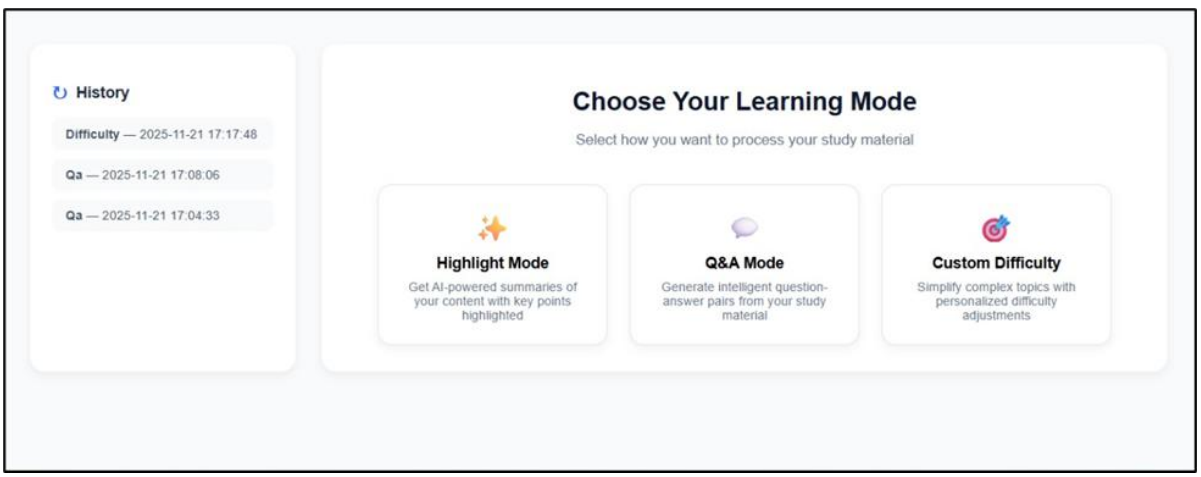

Figure 2: Mode Selection

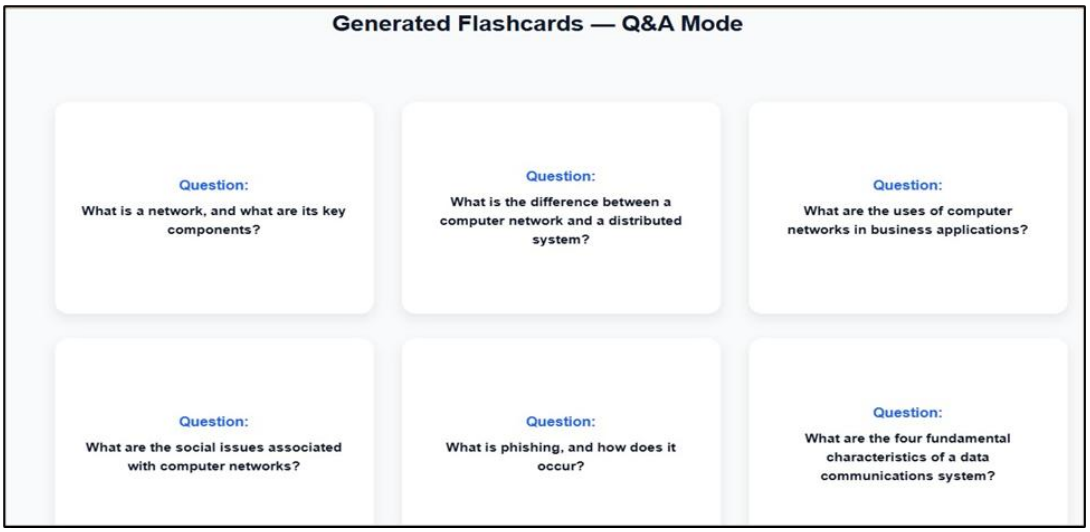

Figure 3: Q & A Mode

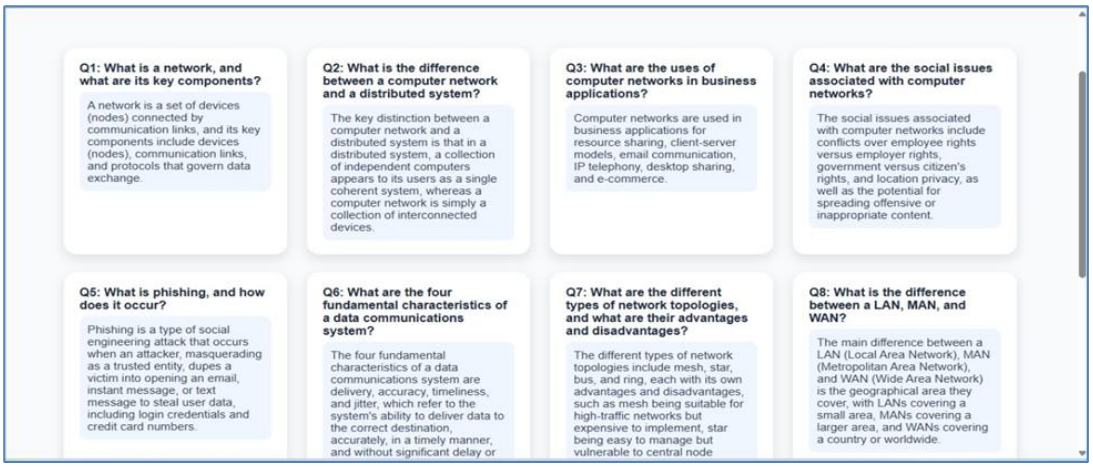

Figure 4: Show all mode

VIII. Performance Evaluation

TABLE 1. Comparision with previous systems

Metric	FlashMe(2023)	Q&A Transformer (2023)	WikiFlash(2021)	Our System
Flashcard accuracy	78%	80%	74%	82%
Concept coverage	60%	65%	67%	80%
Processing Speed (Avg. seconds)	40 seconds	35 seconds	45 seconds	25-30 seconds
Computational Efficiency	High GPU needed	GPU required	GPU required	Runs on normal laptop
Image-to-Text Understanding Score	No images	No images	No images	Takes images (Blip caption s+ OCR)

- The flashcard accuracy shows a notable improvement, with the current system reaching 82%, surpassing FlashMe (78%), Q&A Transformer (80%), and WikiFlash (74%).
- Concept coverage is significantly higher at 80%, indicating better extraction of meaningful ideas from study material compared to earlier models, which range between 60–67%.
- Processing speed is faster, averaging 25–30 seconds, while previous systems take 35–45 seconds, demonstrating reduced computational latency and quicker flashcard generation.
- The system operates efficiently on a normal laptop, unlike the earlier models that demand GPU-level hardware, improving usability in low-resource environments.
- Image handling capability is an added advantage; earlier systems process only text, while the current system supports image-to-text understanding through BLIP captioning and OCR, enabling richer and more diverse flashcard generation.

IX. Conclusion

The AI-powered Flashcard Generator was developed to support smarter and more efficient learning by automatically extracting key information from text, PDFs, and images. Learning effectiveness is strengthened through highlight, Q&A, and difficulty-based modes, while OCR and image-captioning capabilities broaden the range of usable study materials. Interactive flashcards and downloadable PDFs further simplify review and reinforce understanding.

References

[1]. Thabet, Baha, Et Al. "Q&A Generation For Flashcards Within A Transformer-Based Framework." Communications In Computer And Information Science, Vol. 1812, Springer, 2023, Pp. 789–806.
[2]. Pikle, Shlok, Et Al. "Flashme: Automatic Flashcard Generation." 2023 International Conference On Advances In Computing, Communication And Applied Informatics (ACCAI), IEEE, 2023, Vol. 2023, Pp. 1–6.
[3]. Bachiri, Et Al. "Optimizing Learning Outcomes And Retention In Moocs With AI-Generated Flashcards." Lecture Notes In Educational Technology, Vol. 358, Smart Learning For A Sustainable Society, Springer Nature, 2023, Pp. 242–247.
[4]. Ghosh. "Interactive Augmented Reality Application Using Animal Flashcards For Education Of Children." Taylor & Francis, 2025
[5]. Cheng, Yuang, Et Al. "Wikiflash: Generating Flashcards From Wikipedia Articles." Lecture Notes In Computer Science, Vol. 12908, Springer, 2021, Pp. 138–149.

.