



# AI for Automated Generation of Presentation Slides from Text Notes

NIVETHA S<sup>1</sup>, KEERTHANA S<sup>2</sup>, HEMASRI S<sup>3</sup>, SUGUMARAN V R<sup>4</sup>

*1, 2, 3 Members - 5th Semester B.E Students, Department of Computer Science and Engineering,  
E.G.S.Pillay Engineering College, Nagapattinam, Tamilnadu, India*  
*4 Professor, Department of Computer Science and Engineering, E.G.S.Pillay Engineering College,  
Nagapattinam, Tamilnadu, India*

**Abstract** — Presentation creation from lengthy textual documents is a time-consuming and repetitive task that requires significant manual effort in extracting important points, organizing content, and designing slides. Existing presentation generation systems often rely on template-based approaches and fail to provide fully automated and meaningful slide structuring. This paper presents an AI-based system for the automated generation of presentation slides from text notes using Natural Language Processing (NLP) techniques. The proposed system accepts textual input in formats such as PDF, DOCX, and TXT, preprocesses the content, extracts important keywords and sentences, and organizes the extracted information into structured slide content. The system automatically generates PowerPoint presentations using Python-based libraries and NLP models.

Experimental results demonstrate that the proposed approach reduces manual effort, improves efficiency, and generates readable and well-structured slides.

The proposed framework has significant potential for future enhancement through advanced AI models, intelligent design generation, and multilingual support.

**Keywords** — Artificial Intelligence, Natural Language Processing, Presentation Automation, Text Summarization, Slide Generation, Machine Learning.

## I. INTRODUCTION

In academic institutions, business organizations, and professional environments, presentations are widely used for communication and knowledge sharing. Preparing presentation slides from large textual documents is often a labor-intensive process that

requires identifying important information, summarizing content, and organizing it into a visually understandable structure. Manual slide preparation consumes considerable time and may result in inconsistent formatting and incomplete content extraction.

Recent advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have enabled the automation of several document-processing tasks, including summarization, keyword extraction, and intelligent content organization. AI-driven systems can process large volumes of unstructured text and transform them into concise and meaningful information suitable for presentation generation.

This paper proposes an AI-based automated slide generation system that converts textual notes into PowerPoint presentations. The system employs NLP techniques for preprocessing, sentence tokenization, keyword extraction, and summarization. The extracted information is automatically organized into slide titles and bullet points, and presentation slides are generated using Python libraries. The proposed system aims to reduce manual effort, improve presentation creation efficiency, and produce structured and meaningful slide content automatically.



## II. PROBLEM STATEMENT

Creating presentation slides from lengthy text notes is a challenging and time-consuming task. Students, researchers, and professionals often spend considerable effort identifying important points and arranging them into an organized presentation format. Existing systems generally depend on predefined templates or manual editing, limiting automation and content quality.

Most currently available tools are unable to accurately extract important information from unstructured text and frequently generate incomplete or poorly structured slides. Additionally, many systems lack intelligent summarization capabilities and fail to maintain logical content flow between slides. Therefore, there is a need for an AI-based automated system capable of generating clear, meaningful, and structured presentation slides directly from textual notes.

## III. OBJECTIVES

The primary objectives of the proposed system are as follows:

1. To develop an AI-based system for automatically generating presentation slides from text input.
2. To apply Natural Language Processing techniques for extracting important information from textual documents.
3. To reduce manual effort and improve efficiency in presentation preparation.
4. To generate structured, readable, and meaningful slide content automatically.
5. To support multiple document formats such as PDF, DOCX, and TXT.
6. To improve content organization through automated summarization and keyword extraction.

## IV. LITERATURE REVIEW (SUMMARY)

Earlier presentation generation systems primarily relied on rule-based and template-based approaches.

These systems generated slides using predefined structures but lacked flexibility and intelligent content extraction capabilities. As a result, generated presentations often contained irrelevant or incomplete information.

Recent research has focused on integrating Artificial Intelligence and Natural Language Processing techniques into automated slide generation systems. Advanced NLP models have improved text summarization, topic identification, and keyword extraction, enabling better content organization and slide quality.

Several modern systems utilize transformer-based architectures and intelligent workflows to produce more accurate and context-aware presentations. However, many existing approaches still face limitations in terms of full automation, scalability, and user-friendly implementation. Most systems also lack support for intelligent design generation and detailed content structuring.

The proposed system addresses these limitations by integrating NLP-based preprocessing, automated summarization, content structuring, and PowerPoint generation into a unified framework.

## V. PROPOSED SYSTEM

The proposed system uses Artificial Intelligence and Natural Language Processing techniques to automate presentation slide generation from textual notes. The system accepts input documents in formats such as PDF, DOCX, and TXT. The input text is preprocessed to remove unwanted characters, stop words, and irrelevant information.

After preprocessing, the system performs sentence tokenization, keyword extraction, and text summarization to identify important content. The extracted information is organized into slide titles and bullet points according to logical topic flow.

The structured content is then converted into PowerPoint presentation slides using Python libraries such as python-pptx. The generated presentation can



be previewed and downloaded by the user. The proposed system reduces manual effort while improving the efficiency and consistency of presentation generation.

## VI. SYSTEM ARCHITECTURE

The architecture of the proposed AI for Automated Generation of Presentation Slides from Text Notes is designed as a modular and sequential processing framework. The system converts unstructured textual data into meaningful and well-organized presentation slides using Artificial Intelligence and Natural Language Processing techniques. The architecture consists of six major modules, namely Input Module, Preprocessing Module, NLP Processing Module, Content Structuring Module, Slide Generation Module, and Output Module. Each module performs a specific operation to ensure efficient processing and accurate slide generation.

### A. Input Module

The Input Module acts as the initial stage of the system architecture. This module is responsible for accepting textual input from the user in different document formats such as PDF, DOCX, and TXT. The uploaded documents may contain raw text, lecture notes, reports, research content, or academic material.

The system validates the uploaded file to ensure format compatibility and readability. Once validation is completed, the textual content is extracted from the document using appropriate Python libraries such as PyPDF2 and python-docx. The extracted text is then forwarded to the preprocessing stage for further analysis.

The Input Module is designed to support flexible and user-friendly interaction while maintaining compatibility with multiple file formats.

### B. Preprocessing Module

The Preprocessing Module performs text cleaning and normalization operations on the extracted content. Since raw textual data often contains unwanted symbols, punctuation marks, extra spaces, and irrelevant information, preprocessing is essential to improve the efficiency of Natural Language Processing operations.

The preprocessing stage includes:

1. Removal of stop words
2. Elimination of special characters and punctuation
3. Tokenization of sentences and words
4. Conversion of text into lowercase
5. Removal of unnecessary whitespace
6. Text normalization and stemming

The module uses NLP libraries such as NLTK and spaCy to perform these operations efficiently. Preprocessing improves the quality of extracted information and ensures accurate summarization and keyword extraction in subsequent stages.

### C. NLP Processing Module

The NLP Processing Module is the core component of the proposed system. This module applies Natural Language Processing techniques to identify important information from the cleaned text data.

The module performs the following operations:

1. Sentence Tokenization  
The input text is divided into meaningful sentences for easier analysis.
2. Keyword Extraction  
Important keywords and phrases are identified using frequency analysis and NLP algorithms.
3. Topic Identification  
The system determines the major topics discussed within the text document.
4. Text Summarization  
Important sentences are selected to generate concise summaries suitable for presentation slides.



## 5. Semantic Analysis

The relationship between words and sentences is analyzed to maintain meaningful content flow.

The NLP Processing Module ensures that only relevant and informative content is selected for slide generation. This significantly reduces unnecessary information and improves presentation quality.

### ***D. Content Structuring Module***

The Content Structuring Module organizes the extracted and summarized information into a presentation-friendly format. This module converts unstructured text into logically arranged slide components such as titles, headings, subheadings, and bullet points.

The module performs the following tasks:

1. Generation of slide titles from identified topics
2. Arrangement of summarized sentences into bullet points
3. Maintenance of logical content flow between slides
4. Limitation of content quantity per slide for readability
5. Categorization of information into separate sections

The system ensures that each slide contains concise and meaningful information to improve audience understanding. Proper structuring also enhances presentation clarity and visual organization.

### **E. Slide Generation Module**

The Slide Generation Module automatically creates PowerPoint presentation slides from the structured content generated in the previous stage. This module uses Python libraries such as python-pptx for presentation creation and formatting.

The module performs several functions including:

1. Creation of presentation slides
2. Insertion of slide titles and bullet points
3. Application of basic formatting styles
4. Adjustment of text alignment and layout
5. Addition of slide numbering and themes

Each slide is generated dynamically based on the extracted content. The module ensures consistency in formatting and reduces manual presentation design effort.

The generated slides maintain readability and professional structure while preserving the original meaning of the input content.

### ***F. Output Module***

The Output Module represents the final stage of the system architecture. This module generates the completed PowerPoint presentation and provides it to the user for preview and download.

The module supports:

1. PowerPoint presentation export (.pptx)
2. Preview of generated slides
3. Storage of generated presentations
4. Download functionality for users

The final output contains automatically generated presentation slides with organized titles, bullet points, and structured information extracted from the input text document.

The Output Module ensures smooth delivery of the generated presentation while improving user accessibility and usability.



## VII. METHODOLOGY

The methodology of the proposed AI for Automated Generation of Presentation Slides from Text Notes follows a systematic and sequential workflow for converting unstructured textual documents into organized PowerPoint presentations. The methodology integrates Artificial Intelligence and Natural Language Processing techniques to automate text analysis, content extraction, summarization, and slide generation. The overall process consists of multiple stages including data input, preprocessing, NLP analysis, content structuring, slide generation, and output delivery.

The detailed methodology of the proposed system is explained below.

### *Step 1: Data Collection and Input Acquisition*

The first step of the methodology involves collecting textual input from the user. The system accepts documents in multiple formats such as PDF, DOCX, and TXT. These documents may contain lecture notes, reports, academic materials, research papers, or textual content intended for presentation creation.

The uploaded files are validated to ensure compatibility and readability. After validation, the system extracts textual content from the document using document-processing libraries such as PyPDF2 and python-docx.

The extracted text serves as the primary input for further processing stages.

### *Step 2: Text Preprocessing*

The extracted text often contains unwanted symbols, punctuation marks, stop words, and formatting inconsistencies that may affect the accuracy of content extraction. Therefore, preprocessing is performed to clean and normalize the text data.

The preprocessing stage includes the following operations:

#### **1. Stop Word Removal**

Commonly occurring words such as “is,” “the,” and “and” are removed since they do not contribute significantly to content understanding.

#### **2. Tokenization**

The text is divided into smaller units such as sentences and words to simplify further analysis.

#### **3. Lowercase Conversion**

All text characters are converted into lowercase to maintain uniformity.

#### **4. Removal of Special Characters**

Symbols, punctuation marks, numbers, and unwanted characters are eliminated.

#### **5. Text Normalization**

Words are normalized using stemming and lemmatization techniques to reduce words to their root forms.

Natural Language Processing libraries such as NLTK and spaCy are used to perform preprocessing efficiently.

### *Step 3: Sentence Analysis and Keyword Extraction*

After preprocessing, the system analyzes the cleaned text to identify important information required for presentation generation.

The NLP Processing stage performs the following operations:

#### **1. Sentence Tokenization**

The cleaned text is segmented into meaningful sentences for analysis.

#### **2. Keyword Extraction**

Important keywords and phrases are extracted using word frequency analysis and NLP-based ranking algorithms.



### 3. Topic Identification

The system identifies major topics discussed in the document by analyzing keyword occurrences and contextual relationships.

### 4. Important Sentence Selection

Sentences containing high-information content are selected based on relevance scoring techniques.

The extracted keywords and sentences form the basis for slide content generation.

#### *Step 4: Text Summarization*

The selected sentences are summarized to generate concise and presentation-friendly content. Since presentations require short and meaningful points rather than lengthy paragraphs, summarization plays a critical role in improving slide readability.

1. The summarization process includes:
2. Identification of high-priority sentences
3. Removal of redundant information
4. Compression of lengthy text into concise points
5. Preservation of important concepts and meanings

The summarized content ensures that generated slides remain informative while avoiding excessive textual information.

#### *Step 5: Content Structuring*

The summarized information is organized into a structured presentation format. The Content Structuring stage arranges the extracted content into titles, headings, and bullet points.

This stage performs the following tasks:

1. Generation of slide titles from identified topics
2. Organization of summarized content into bullet points
3. Logical sequencing of slides

4. Limiting the amount of text per slide
5. Grouping related information together

The module ensures that the presentation follows a meaningful and understandable flow between slides.

#### *Step 6: Automatic Slide Generation*

After content structuring, the system automatically creates PowerPoint presentation slides using Python libraries such as python-pptx.

The Slide Generation stage includes:

##### 1. Slide Creation

New slides are dynamically generated for each identified topic.

##### 2. Title Insertion

Generated slide titles are inserted into the presentation layout.

##### 3. Bullet Point Addition

Summarized content is added as bullet points under respective headings.

##### 4. Formatting and Layout

Basic formatting such as font size, alignment, spacing, and slide themes are applied automatically.

##### 5. Presentation Organization

Slides are arranged sequentially to maintain presentation consistency.

This automation significantly reduces manual slide preparation effort.

#### *Step 7: Presentation Output Generation*

The final stage of the methodology generates the completed PowerPoint presentation file. The generated slides are combined into a structured presentation and exported in PPTX format.



The Output Generation stage supports:

1. Presentation preview
2. PPTX file generation
3. Download functionality
4. Storage of generated presentations

Users can review, edit, and download the generated presentation according to their requirements.

## VIII. RESULTS AND DISCUSSION

The proposed AI for Automated Generation of Presentation Slides from Text Notes system was implemented and evaluated using different categories of textual documents including academic notes, technical reports, seminar materials, and research content. The performance of the system was analyzed based on content extraction accuracy, slide generation quality, readability, processing efficiency, and presentation structure.

The system successfully converted unstructured textual input into organized PowerPoint presentation slides using Natural Language Processing techniques. The preprocessing module effectively cleaned the textual data by removing stop words, punctuation marks, and irrelevant symbols, thereby improving the efficiency of subsequent NLP operations. Sentence tokenization and keyword extraction modules accurately identified important information from the input documents.

The summarization process generated concise and meaningful content suitable for presentation slides. Important topics and key points were extracted and organized into headings and bullet points with proper logical flow. The generated slides were readable, well-structured, and suitable for academic and professional presentation purposes.

The system was tested with multiple document formats including PDF, DOCX, and TXT files. Experimental results demonstrated that the proposed framework could successfully process different types of input documents without significant performance degradation. The generated presentations maintained consistency in slide formatting and content organization.

The automatic slide generation module created PowerPoint slides dynamically using the extracted summarized content. Basic formatting features such as title alignment, bullet arrangement, font styling, and slide sequencing were implemented successfully using the python-pptx library. The generated slides reduced the need for manual editing and significantly minimized presentation preparation time.

The performance of the proposed system was evaluated using the following parameters:

1. Content Extraction Accuracy
2. Quality of Generated Summaries
3. Slide Structure and Readability
4. Processing Speed
5. Presentation Consistency

The experimental observations indicated that the system achieved satisfactory results in extracting meaningful information and generating structured presentation content. The generated slides effectively represented the major concepts present in the original text document.

Furthermore, the proposed system demonstrated the practical applicability of Artificial Intelligence and Natural Language Processing techniques in automating document-to-presentation conversion tasks. The automation process improved efficiency and reduced the workload associated with manual slide preparation.

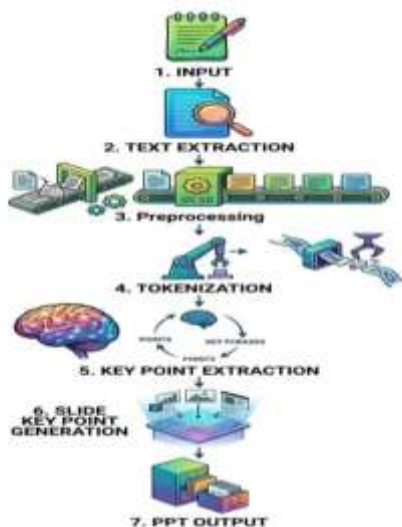
However, certain limitations were identified during testing. In some cases, highly complex or lengthy documents produced partially incomplete summaries due to limitations in the summarization process. Additionally, the current implementation provides only basic slide formatting and lacks advanced visual design features such as automatic image generation, smart animations, and adaptive themes.

Despite these limitations, the overall performance of the proposed system was found to be effective for automated presentation generation. The system provides a strong foundation for future enhancements involving transformer-based NLP models, intelligent



slide design systems, and AI-driven visual content generation.

The results confirm that the proposed AI-based framework can significantly improve productivity for students, researchers, educators, and professionals by automating the process of presentation slide generation from textual notes.



## IX. ADVANTAGES

### 1. Reduction of Manual Effort

The system automatically converts textual documents into presentation slides, thereby reducing the manual effort required for slide preparation.

### 2. Time Efficiency

The proposed framework significantly reduces the time required for creating PowerPoint presentations from lengthy text documents.

### 3. Automatic Content Extraction

Natural Language Processing techniques are used to automatically extract important keywords, topics, and meaningful sentences from input text.

## 4. Structured Slide Organization

The generated slides contain properly organized titles, headings, and bullet points with logical content flow, improving readability and presentation quality.

## 5. Support for Multiple Document Formats

The system supports different input formats such as PDF, DOCX, and TXT, increasing flexibility and usability.

## 6. Improved Productivity

The automation of presentation generation improves productivity for students, researchers, educators, and professionals by minimizing repetitive manual tasks.

## X. LIMITATIONS

Despite the effectiveness of the proposed AI for Automated Generation of Presentation Slides from Text Notes system, certain limitations exist in the current implementation.

### 1. Limited Content Detailing

The generated slides may sometimes contain brief or incomplete information, particularly for highly technical or lengthy documents.

### 2. Basic Slide Design

The current system supports only basic slide formatting and does not include advanced visual design features such as animations, smart templates, or adaptive layouts.

### 3. Dependency on Input Quality

The accuracy of generated presentations depends heavily on the quality and clarity of the input text document. Poorly structured input may affect content extraction performance.

### 4. Limited Understanding of Context

The system primarily relies on NLP-based summarization techniques and may not fully



understand complex contextual relationships within the document.

### 5. Absence of Automatic Image Generation

The present implementation focuses mainly on textual content generation and does not automatically include relevant images, charts, or diagrams in slides.

### 6. Limited Multilingual Support

The system is currently optimized for English text documents and provides limited support for multilingual presentation generation.

## XI. FUTURE SCOPE

The proposed AI for Automated Generation of Presentation Slides from Text Notes system can be further enhanced by integrating advanced Artificial Intelligence and Natural Language Processing techniques to improve automation accuracy and presentation quality. Future improvements may include the implementation of transformer-based NLP models such as BERT and GPT for better text summarization, topic identification, and contextual understanding. The system can also be extended to automatically generate relevant images, charts, and diagrams to improve the visual quality of presentation slides. Advanced slide templates, smart layouts, and animation features may be incorporated to create more professional and attractive presentations. In addition, multilingual support can be implemented to allow presentation generation from documents written in different languages. The framework can also be integrated with cloud-based and web platforms to provide online accessibility, real-time processing, and collaborative presentation editing. Furthermore, integration with platforms such as Google Slides and Microsoft PowerPoint can improve usability and presentation management. These future enhancements can significantly increase the effectiveness, scalability, and practical applicability of the proposed system in academic and professional environments.

## XII. CONCLUSION

This paper presented an AI for Automated Generation of Presentation Slides from Text Notes system that utilizes Artificial Intelligence and Natural Language Processing techniques to automatically convert textual documents into structured PowerPoint presentations. The proposed system successfully performs text preprocessing, keyword extraction, summarization, content structuring, and automatic slide generation to reduce the manual effort involved in presentation preparation.

The implementation demonstrates that the system can effectively extract important information from unstructured text documents and organize the content into meaningful slide titles and bullet points. The generated presentations are readable, structured, and suitable for academic as well as professional use. The proposed framework significantly improves time efficiency and productivity by automating the process of presentation creation.

Although the current implementation provides only basic formatting and summarization capabilities, the system establishes a strong foundation for future intelligent presentation generation frameworks. Future enhancements involving advanced NLP models, automatic image generation, multilingual support, and cloud integration can further improve the quality and functionality of the system. Overall, the proposed system demonstrates the practical applicability of AI-driven automation in simplifying presentation slide generation from textual notes.

## REFERENCES

- [1] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. Stanford, CA, USA: Stanford University, 2024.
- [2] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. Sebastopol, CA, USA: O'Reilly Media, 2023.
- [3] "python-pptx Documentation," Python Software Foundation. [Online]. Available: <https://python-pptx.readthedocs.io/>



- [4] “Natural Language Toolkit (NLTK) Documentation,” NLTK Project. [Online]. Available: <https://www.nltk.org/>
- [5] “spaCy Documentation,” Explosion AI. [Online]. Available: <https://spacy.io/>.
- [6] A. Vaswani *et al.*, “Attention is all you need,” in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, 2017, pp. 5998–6008.
- [7] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–4186.
- [8] Y. Liu and M. Lapata, “Text summarization with pretrained encoders,” in *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong, China, 2019, pp. 3730–3740.
- [9] A. Kumar and R. Sharma, “AI-Based Automated Presentation Generation Using Natural Language Processing Techniques,” in *Proc. IEEE International Conference on Artificial Intelligence and Data Engineering (AIDE)*, Chennai, India, 2024, pp. 112–118.
- [10] P. Singh and V. Rao, “Automated slide generation from text documents using machine learning and NLP approaches,” *IEEE Access*, vol. 12, pp. 44567–44579, 2025.