International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 14; Issue: 8; August- 2025

SMART TEXT SNAP SUMMARIZER: "REAL-TIME SUMMARIZATION OF CAPTURED TEXT FROM IMAGE AND SCREENSHOTS USING AI"

Vinay kumar N P
PG, Student
The Oxford College of Engineering,
Bommanahalli,
Bengaluru- 560068
vinaykumarnpmca2025@gmail.com

Sowmya J
Assistant Professor
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
sowmyaj@theoxford.edu

ABSTRACT

In today's world of information overload, it's challenging to find useful information within vast amounts of text. The project is the development of an automated document summarizer based on deep learning. It incorporates neural networks structures to produce clean and rational summaries out of bulky materials. It uses a contemporary algorithm that uses sequence-to-sequence approach, attention mechanism and transformer based architectures to get the phrase into context and give meaningful summaries. Evaluation of this technique shows that it has an impressive effect on the quality, coherency, and relevance of summaries in comparison with the more conventional methods of extraction. The technology may be of use in research, aggregation of news, analysis of legal documents, and other studies where fast access to information is needed.

Keywords:

Auto-Summarization, Deep learning, Sequence-to-sequence, Transformer, Attention mechanism, Document mining, Text mining, Natural language processing (NLP).

INTRODUCTION

It is difficult for People and businesses are finding it hard to comprehend and interpret bulks of text as the amount of information and data are rapidly increasing in most spheres. Reading through some literature, including research papers, news articles, legal documents or company reports is a time-consuming process and very error-prone as far as capturing relevant information is concerned. Automated text summary has become an important aspect of research in natural language processing (NLP) since it leads to the formulation of short, clear and useful summaries of long texts.

Most conventional approaches to summarizing use fixed heuristics, frequency-based measures or keyword extraction. Rules or statistics are often used in such approaches. While they can

produce simple summaries, they often fail to capture the full meaning, context, and connections within the material. As a result, the summaries may be inaccurate or incomplete.

Deep learning methods have gained significant interest as a solution to these limitations due to their ability to represent complex language structures and semantic links within text. The field has progressed because of models like sequence-to-sequence (Seq2Seq) networks, attention mechanisms, and transformer designs like BERT and GPT. These models allow computers to understand sentence context, identify important information, and create summaries that feel human-written.

LITERATURE SURVEY:

1.EXISTING WORK

Sangita Pokhrel, Swathi Ganesan, Tasnim Akther, and Lakmali Karunarathne's research paper, "Building Specific Chatbots for Document Summarization and Question Answering employing Large Language Models with a a novel framework for creating customized chatbots that use large language models (LLMs) for document summarization and question answering. The framework incorporates cutting-edge tools like Streamlit and LangChain to effectively manage information overload by gleaning important insights from long documents. The study examines the framework's design, implementation, and practical uses, emphasizing how it can boost output and facilitate effective information retrieval. The paper illustrates its practical significance by showing developers how to create end-to-end applications for document summarization and question answering by offering comprehensive guidance. Challenges in online education are addressed in the research paper "State-of-the-Art Approach to e-Learning with Cutting-Edge NLP Transformers: Implementing Text Summarization, Question and Distractor Generation, Question Answering" by S. Patil, Lokshana Chavan, Janhvi Mukane, D. Vora, and V. Chitre. To improve e-learning, the study uses linguistic resources like WordNet, ConceptNet, and Sense2Vec in addition to sophisticated NLP transformer models like T5, DistilBERT, and DistilBART. The suggested system generates multiple-choice questions (MCQs), fill-in-the-blanks, and single-word answers in addition to enabling users to upload e-books and receive text summaries. Users are able to assess their answers, making it a complete learning and assessment tool. The efficiency of online learning is increased and content accessibility is

streamlined with this automated method of question generation and summarization.

2.PROPOSED SYSTEM

1. Automated Text Extraction and Preprocessing:

The system will automatically extract text from different document formats, including scanned PDFs, images, and handwritten notes.

- Preprocessing steps like noise removal, OCR (Optical Character Recognition), tokenization, and normalization will be applied to prepare the text for summarization.

- This keeps the input clean and structured, which improves the efficiency of the summarization process.

2. **Deep Learning-Based Summarization Engine** :

- The main engine uses deep learning models, like Transformer-based models such as BERT, GPT, or T5, for summarization.

- It creates concise, relevant summaries while maintaining the main ideas of the document.

- The system supports both extractive summarization, which selects key sentences, and abstractive summarization, which generates new sentences.

3. Multi-Format and Multi-Language Support :

- It can handle documents in various formats, like PDF, DOCX, JPEG, and PNG, and convert them into readable text.

- The system offers potential multilingual summarization for documents in different languages using pre-trained multilingual models.

METHODOLOGY

An automated pipeline for smart text snap summarization was developed as a multi-step system integrating: OCR, NLP preprocessing, and summarization techniques. The process begins when input is captured in the form of an image, and images are accessed, and document or free text provided by the taking a picture of the document and text storage in a common image format. Text is distilling out from the image using the pytesseract library developed in Python by Google and based on Tesseract OCR engine, and distilling out from an image is called Optical Character Recognition, or OCR. After the image has been converted into machine readable text, the text is then preprocessed with various steps involving noise reduction, lowering case, stop-word removal, tokenization, and lemmatization, performed by Natural Language Toolkit (NLTK) for example, or spaCy, then the text quality and consistency will be improved and more suitable as an input to the next layer of summation.

EXPERIMENTAL RESULTS

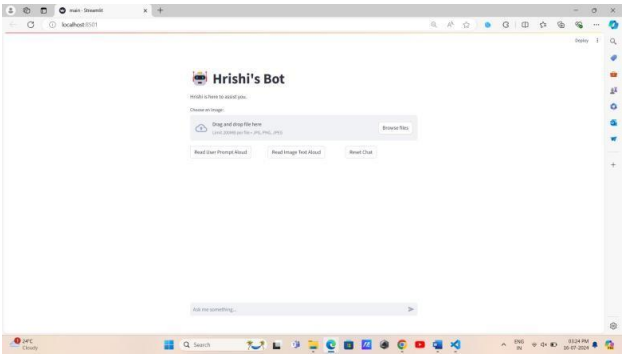

Fig 1 User Interface Page using Streamlit component

Streamlit forms the basis for the design of the user interface for the Image-to-Text Summarizer System, which allows for an easy and simple environment for users to interact with the program. Through this interface, users are able to upload images of any type like screenshots of research, scanned paperwork and even illegible scribbles at just a few clicks. After the uploading of the image, the interface links to backend features, the summarization module and OCR which extract text all to give short summaries. The extracted text and summaries were placed in a text interface that is well organized and constructive allowing the user to see the name and extracted text simultaneously. A potential positive of streamlining the interface in Streamlit is that aspects like download buttons of the summaries, the visual display of the results and the ability to upload images with which the user can interact, are much easier.

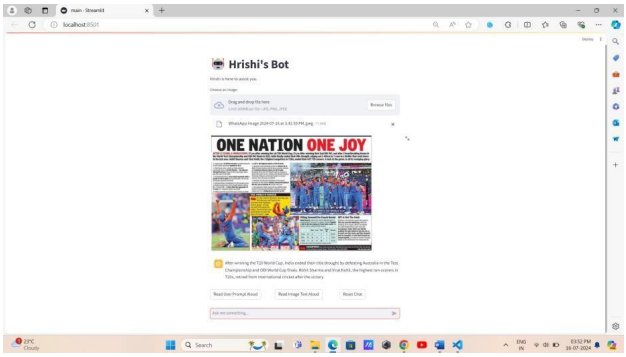

Fig 2 Image Upload its Summary Generated

Once the images are uploaded to the system using the Streamlit interface, the OCR module is used to extract data contained in the images translated into text form. After the text has been extracted, it will be summarized into a simple easy to read summary of the text by the summarization module. This means that the user can read Fragmented summary of the text thereafter in the interface to ensure original concepts are obvious in the image. The mechanism of turning such unstructured visual data into consumable information enables users to attain the filled in ideas in few moments of time, with no reading of the entire image of the information. The system is simple and effective since the image uploading, text extraction and summary process is performed
cumulatively.

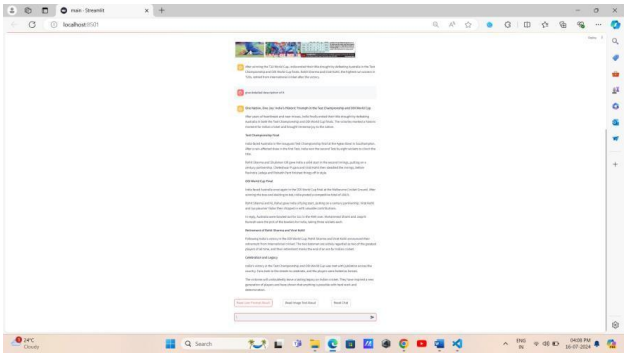

Fig 3 Reading the AI response in speech

After the AI module has processed a user query, and produced a response from the queried text, the interface can read the response using a text-to-speech feature. This feature provides clear sounding speech output from the AI-generated text allowing users to listen to the information rather than read it. This is not only more accommodating for users who may be visually impaired, but it is also more convenient for someone who may want to consume certain information auditorily as opposed to visually. By providing speech output alongside an AI powered response to the user query, the interface provides a more interactive, engaging and usable experience. This allows users to receive and comprehend information, all at the same time.

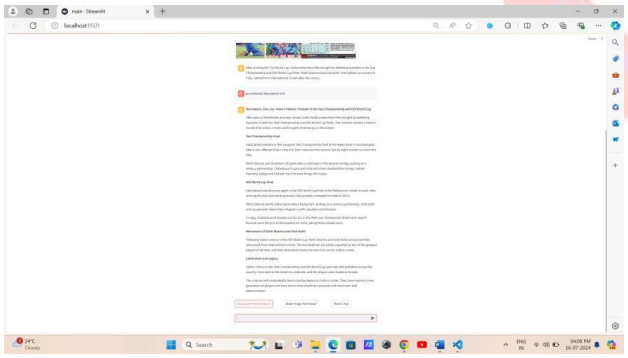

Fig 4 Reading the AI response in speech

After the AI module processes the user query and generates a response based on the parsed text, the system has the option to have a text-to-speech function read the response, converting the AI-generated text to spoken language. This text-to-speech function produces clear and natural speech, so the users can listen to the information rather than read it. This feature gives visually impaired users greater access and provides no additional benefit for those users who naturally want auditory forms of feedback. The combination of AI-driven responses to user queries and spoken output strengthens the interactivity, engagement, and friendliness of the interface and the user experience, as the user can receive and understand information.

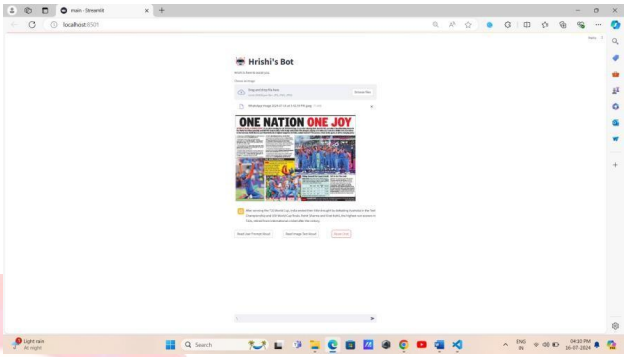

Fig 5 Clearing previous interaction

The system presents the possibility to erase former actions in order to refresh the interface so that the users can perform new actions without retain the data of the former uploads or queries. This operation deletes the images already uploaded, the text extracted, the text summary created, the text responses generated by AI and the speech generated. Questioners and respondents now have the opportunity to begin a new interface with no links to the past. The rationale in having this ability is the undesirability of subsequent confusion in case one were to re-enter, or out of ignorance, rediscover an old task which might have confused interference intruded or overlapped with a new task of previous peregrination. Therefore, this authorizes a more direct use of: Clearing past activity also enhances system performance by liberating found and usable memory and resources thus creating the

optimum performance of the application at a time.

CONCLUSION

A significant advancement in artificial intelligence and document processing is the establishment of this framework for processing images, generating summaries, and answering user questions. The combination of several high-end technologies, including Google Generative AI (Gemini AI) models, Optical Character Recognition (OCR) using Pytesseract, and an extremely easy to use interface built with Streamlit, can be useful together. All of the components work together to provide a productive, flexible, and easy-to-use system that is able to meet a wide variety of needs.

The use of OCR is essential due to the ease, accuracy, and speed of automatically extracting text data from a number of image types, including scanned documents, handwritten notes, screenshots and other image files. The system saves time, allows you to do less manual transcribing and reduces errors by making the process entirely automated! High quality AI-based model, means that the outputs for example question/answer, and summaries are from advanced AI models that have been fine-tuned on a wide variety of datasets, meaning that their answers and summaries are relevant and contextually similar, guaranteeing the quality of the information from the visual data.

REFERENCES

[1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[2] Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016**).** SQuAD: 100,000+ Questions for Machine Comprehension of Text. arXiv preprint arXiv:1606.05250.

[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[4] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. OpenAI Blog.