An AI-Based Storyboard Generation System with NLP and Image Retrieval for Cognitive Support in Special Education

Nisha Wandile Kimmatkar[1], Rutvik Mahale[2], Vismaya Shino[3], Parth Borole[4], Tanvee Kulkarni[5]
[1,2,3,4,5]*JSPMS's Rajarshi Shahu College Of Engineering, Tathawade*

***Abstract*—Even with continuous support from teachers and caregivers, learning and adapting to new technologies can be especially difficult for people with intellectual disabilities. Traditional teaching methods rarely accommodate their needs as learners, thus reducing effectiveness and interest. Neither the learners nor the educators are at fault; instead, these limitations point to a greater need for better and more inclusive teaching practices. For this purpose, the present study proposes a mind-mapping storyboard system that transforms textual information into simple, engaging pictures. For transforming written texts into intuitive visual elements, the proposed system aims at providing an easy and supportive learning environment. This will help in maintaining student attention, enhancing their comprehension, and memory retention of learning material. Ultimately, the aim is to introduce a tool that will make learning both more enjoyable and effective for strengthening educational results, as the current solution fully corresponds to the specific cognitive processing styles of these learners. The project also aims to integrate this innovative technique in order to bridge the disconnect between the traditional educational methods and the needs of intellectually disabled people and to support their academic development and adaptation to the digital learning environment.**

I. INTRODUCTION

Individuals with intellectual or developmental disabilities commonly face difficulties in processing, comprehending, and remembering text-based information. Traditional teaching methods-written notes, verbal descriptions, or static materials displayed in the classroom-do not fully meet their cognitive strengths in processing information. This can result in lower levels of comprehension, scant enthusiasm, and problems learning new information. Such challenges do not arise from a lack of ability to learn but rather as an outcome of the mismatch between traditional teaching formats and the way these learners process the information. As educational settings continue to implement more digital tools and AI, so too have opportunities arisen regarding the design for accessible learning systems that cater to a wide range of cognitively diverse users. AI will not replace human creativity or instruction but can act as an enabling medium to extend how information is presented and perceived. Applied with responsibility, it has the potential to simplify complex ideas and present them in formats that resonate more naturally with visual learners.

The proposed system attempts to fill this gap by converting user-provided text into a mind-mapping storyboard that consists of clear, intuitive, and visually meaningful images. By incorporating Natural Language Processing, NLP, for keyword extraction, and integrating cloud-based image retrieval, the system effectively converts textual content into structured visual representations more easily comprehended by intellectually disabled individuals. This is beneficial in terms of enhancing understanding through visual storytelling, maintaining learner attention, and ultimately enhancing retention.

This focuses on the core ideals of ease, simplicity, and accessibility that underlie the larger objective of designing technologies to adapt to the learner rather than forcing the learner to adapt to the technology. In the end, development of this storyboard-based system represents a step toward more equitable and effective digital learning environments for individuals with intellectual disabilities.

II. LITERATURE REVIEW

Paper 1: "Advancements in Text-to-Image Generation through Generative AI" - (2024)

This work discusses recent generative models, especially GANs, which are composed of a generator and discriminator working in opposition to generate good quality visual outputs. The paper explained the importance of the selection of the dataset and fine-tuning; adapting a pre-trained model to domain-specific datasets increased the accuracy of the image significantly. Another important mechanism mentioned was feedback, which facilitated a continuous improvement in the performance of the model.

However, the authors highlight ongoing problems with textual prompt ambiguity, mode collapse, and irregularities in the quality of images generated that get in the way of generating visuals reliably. These limitations underpin why generative models can struggle when required to provide simple, predictable visuals suitable for cognitively challenged learners.

Paper 2: "Keyword Extraction: A Modern Perspective" (2022)

This paper provides a general overview of various approaches to keyword extraction, from the statistical approaches of TF-IDF to more recent neural network models, which can even learn deep contextual patterns in text. The authors note that deep learning approaches increase keyword relevance but at high computational cost.

Key challenges include ambiguity in defining important terms, dependency on high-quality datasets, and the possibility of missing contextual keywords not explicitly present in the text. The findings substantiate the need for robust NLP pipelines in applications that rely on accurate concept extraction, such as the storyboard system designed in this project.

Paper 3: "Text-to-Image Generation using Generative AI"

This paper investigates text-to-image workflows that involve Text CNN-based encoders for enhancing the model's ability in capturing local semantic features. Another important aspect explored is feature-fusion techniques that show how the integration of textual semantics with visual features can help to enhance generated images.

Despite these advances, the study identifies challenges in maintaining semantic consistency, handling structural ambiguity, and generalizing across varied inputs of text. High computational overhead further restricts practical deployments. These somewhat confirm that purely generative approaches might not always be apt for accessibility-oriented tools needing consistent and predictable visuals.

Paper 4: "Integrating AI-Generative Tools in Web Design Education" (2023) This mixed-methods study investigates how the integration of AI-based content generation tools into web design courses influences creative processes. The findings, based on surveys, project evaluations, and reflective feedback, indicate that AI can augment creativity by serving as a supportive tool, not a replacement for human input. It also reports concerns about authenticity, copyright issues, and variable levels of student adaptability. Not directly related to intellectual disability, these findings nonetheless reinforce the importance of designing AI systems that support—and do not dominate—the creative or educational process. Paper 5: "An Analysis on the Use of Image Design with Generative AI Technologies" This research is focused on AI-generated imagery within professional design workflows. It outlines the strengths of neural methods in producing visually compelling outputs while identifying concerns around bias, loss of artistic authenticity, and overdependence on AI tools. As part of the discussion, the paper points out that the complexity of generative models often obscures user understanding and, therefore, is at odds with transparency and trust. These observations stress the importance of simple, intuitive AI approaches, especially when designed for special-needs learners.

The education of students with intellectual disabilities has long been a priority, as traditional pedagogical methods cannot typically meet the cognitive and perceptual demands that come with a disability. Evidence from studies indicates that numerous students experience significant difficulties in understanding learning materials with high levels of abstraction or a preponderance of text, hence calling for the introduction of educational methods better suited to learning styles. This has inspired movement toward more visually oriented and technology-assisted approaches.

A recurring theme in recent studies is how visual learning tools such as diagrams, images, videos, and mind maps greatly enhance comprehension and retention among learners with cognitive impairments.

These tools enable simplification of complex ideas through the translation of textual information into concrete and easily interpretable visuals, which in turn enhances interrelationships of concepts. Mind mapping has received particular attention in regard to the breakdown of information into structured branches, adding an interactive element to learning and making the process more cognitively approachable.
Concurrently with the improvements in accessibility research, there have also been important developments pertaining to AI-generated content, particularly in text-to-image synthesis, keyword extraction, and automatic visual representation. The works on generative models underpin that these models are able to generate meaningful graphics from a verbal description; however, the ambiguity of text interpretation, instability of models, and high computational burden are still in the way of wide implementation. Research into keyword extraction further shows how statistical, linguistic, and deep learning methods together can extract key ideas from user-generated text.
Other related literature discusses how AI could promote education in general. These studies have also recognized the potential contribution of AI beyond automating information processing towards building more inclusive classes, in particular for learners who benefit from alternative formats of content delivery. Various researchers point to issues such as limitations of datasets, variations in the accuracy of models, and needs for customization when considering the use of AI-based systems for special-needs populations. In general, from the reviewed studies, it can be seen that AI-based generation of visual content is one of the fast-developing areas; however, its application to the field of educating intellectually disabled individuals is rather limited. The findings collectively support the need for solutions that combine accessible interface design, reliable NLP techniques, and meaningful visual representation. This forms the basis for the present work, which integrates these elements into a mind-mapping storyboard system focused on improving learning outcomes among intellectually disabled learners.

III. METHODOLOGY

The system is proposed to generate a mind-mapping storyboard by transforming the text fed by a user into visually meaningful images. The methodology is divided into several stages: from the acquisition of input to text analysis, image retrieval, data handling, generation of visualization, and user-centric evaluation. The following section will elaborate on each stage in detail.

A. Input Acquisition

The first step is the capturing of user input via a simplified web interface that is specially designed for cognitively diverse learners. The interface includes a minimalistic text-entry form where users or caregivers can enter any text that needs to be converted into a visual storyboard. The layout is intentionally uncluttered, with only vital elements visible to avoid cognitive overload. In this system, even those who have limited experience with technology can easily input text without getting confused. Once submitted, this text automatically goes to the backend for further processing. Such a design guarantees accessibility, reduces the degree of interaction complexity, and makes the participation of learners possible, either on their own or requiring only minor assistance.

B. Natural Language Processing and Text Preprocessing

In the second step, the system uses Natural Language Processing to extract meaningful information from the user's text. Preprocessing of the text is carried out by tokenization, removal of stop words, and tagging the part of speech using known NLP libraries such as spaCy and NLTK. These are preparatory steps in which the text is decomposed into manageable linguistic elements that would support deeper semantic processing.
Following that, keyword extraction algorithms pinpoint the core concepts, entities, and actions described in the input. Models like BERT, or other transformer-based architectures, can be integrated to enhance contextual understanding-so that the keywords extracted really convey what the user wants to express. This blend of linguistic and contextual analysis forms the basis for accurate visual mapping in subsequent stages.

C. Image Retrieval Using Open-Source APIs
The system then uses the extracted keywords to fetch corresponding images using open APIs like Unsplash, Pexels, and Google Images. Each keyword is sent to multiple API endpoints so several visually relevant results could be fetched. These images undergo filtering for content that may be inappropriate, too complex, or not suitable for learners with ID.
Other ranking algorithms assess relevance by keyword-image matching score, clarity, simplicity, color contrast, and recognizability. The aim is that every selected image is not only a literal representation of the keyword, but also cognitively friendly for the target audience. This stage will turn abstract words into concrete and visually intuitive representations.

D. Backend Processing and System Integration
The system backend is based on Node.js, acting as a central controller for executing NLP, communicating with the API, and routing data. Once the text input from the frontend is obtained, the backend starts up the NLP pipeline, retrieves keyword outputs, makes image retrieval calls, and compiles the resulting set of images.
A NoSQL database, like MongoDB, stores extracted keywords, past inputs, user preferences, and the URLs of retrieved images. This database allows the system to reduce use of the API by ensuring quick access to frequently used visual content, thus enhancing performance. Node.js also provides a secure processing environment for user data, efficiently handling multiple simultaneous requests. The architecture should be lightweight and scalable, optimized for real-time interaction.

E. Storyboard Generation and Mind-Mapping Visualization
Once relevant images have been retrieved and filtered, the system formats them as a coherent mind-mapping storyboard. The layout will visually organize concepts by placing the central idea at the core and placing associated keywords as branches around it. Each branch shows an image with its corresponding keyword in support of conceptual association.
The storyboard design focuses on high readability, clean spacing, and minimal visual clutter. The use of either cartoon-style or simple photographic images ensures that learners can easily interpret the visual elements. This stage bridges the gap between textual understanding and visual cognition, producing a learning aid that enhances memory retention and comprehension. F. Usability evaluation and system testing A usability evaluation was performed with a target group of individuals with intellectual disabilities and their caregivers to ensure the system would meet the needs of its desired users. Participants used the system to give real-world feedback on clarity, ease of use, navigation effort, and comprehension improvements. Metrics such as keyword extraction accuracy, relevance of retrieved images, response time, and user satisfaction are documented and analyzed. Comparisons may also be made between interface variants or keyword-generation strategies through A/B testing. Insights obtained through these evaluations underpin iterative refinement of the system, ensuring ongoing improvements in terms of accessibility and performance. G. Deployment and Maintenance The final system is deployed on a reliable hosting platform that ensures it is always accessible. Continuous maintenance procedures are implemented to update image sources, improve the quality of NLP models, optimize loading times, and integrate additional features per user feedback. Regular assessment serves to ensure the platform remains relevant, accurate, and tuned to the changing needs of learners.

IV. IMPLEMENTATION

Text Processing
Each sentence entered by the user is transformed by the system's NLP module into a brief list of significant keywords that can subsequently be linked to pertinent images. This system employs a lightweight, browser-based natural language processor (window.nlp) that performs quick text breakdown appropriate for real-time educational applications, in contrast to computationally demanding NLP frameworks. Keyword selection, linguistic category extraction, and text preprocessing make up the NLP pipeline. Below is a detailed description of each step.
A) Tokenization
When a sentence S is passed into window.nlp(sentence), it is immediately segmented into smaller units called tokens. Tokenization is represented as:

$$S = \{w1, w2, w3, \ldots, wn\}$$

Where each wi is an individual word or punctuation mark. is internally processed into:
["The", "boy", "happily", "plays", "in", "the", "park"]

B) Lowercasing and cleaning
To ensure consistent processing, the text is normalized. All tokens are converted to lowercase:

$$wi = \text{lowercase}(wi)$$

And unnecessary whitespace or stray characters are cleaned:
S′=clean(S)

C) Implicit Stop-Word Removal
Although the system does not manually remove stop-words, the NLP library naturally filters them out during category extraction. If SW represents the set of common stop-words, the effective set becomes:

$$T = S' - SW$$

Words like "the," "is," and "in" are ignored because they do not contribute meaningful visual information.
So:
"The boy is playing in the park"
effectively becomes:
["boy", "playing", "park"]

D) Lemmatization
Lemmatization reduces words to their simplest base form:

$$\text{lemma}(wi) = wi$$

Even though your code does not explicitly call a lemmatizer, the NLP engine automatically handles common variations.
Examples include:

"running" → "run"
"stories" → "story"

This enhances the consistency of extracted keywords and helps avoid duplicate or redundant forms.

V. LINGUISTIC ANALYSIS

Once preprocessing is completed, the sentence is analyzed to identify the words that carry real meaning. The system uses built-in NLP functions to extract nouns, verbs, and adjectives from each sentence.
a) Part-of-Speech (POS) Tagging
Each token is assigned a grammatical role:

$$\text{POS}(wi) \in \{\text{noun}, \text{verb}, \text{adjective}, \dots\}$$

This tagging process allows the system to separate objects, actions, and descriptive elements from the rest of the sentence.

b) Noun-Phrase Chunking
Although your implementation extracts only the raw nouns, verbs, and adjectives, the NLP engine is capable of identifying multi-word noun phrases:
For example:
"Little brown dog"
"School playground"

These groupings are useful because many ideas are expressed across multiple words, and they help improve contextual understanding if expanded later.

c) Named Entity Identification (Implicit)
The NLP library can also internally detect named entities such as:
Places
Personal names
Events
Organizations
Mathematically represented as:
This helps the system recognize real-world concepts that may require specific imagery.
Keyword Extraction

A) Keyword Set Construction
Let:
N= set of nouns
V= set of verbs
A= set of adjectives
Then:

$$K = N \cup V \cup A$$

This ensures the final keywords represent the main objects, actions, and descriptions within the sentence.

B) Selecting the Top Three Keywords
To maintain clarity and reduce complexity the system limits the keyword list to the first three meaningful words:

$$K' = \{K1, K2, K3\}$$

C) Creating the Query String
The selected keywords are then combined into a single query string:

$$Q = K1' + "" + K2' + "" + K3'$$

VI. IMAGE RETRIEVAL MODULE

The system's subsequent phase concentrates on obtaining visually relevant images through external APIs after the NLP module generates a refined set of keywords. The Unsplash API is used in the suggested method because of its excellent, varied, and license-friendly image library. However, a CLIP-based semantic ranking model is used to introduce a secondary filtering mechanism because raw API search results might not always be the best match for the user's intent. This combination guarantees that the user will see images that are both aesthetically pleasing and semantically consistent with the extracted keywords' meanings.

The keywords generated from the NLP module are combined to form the final search query Q, represented as:

$$Q = K1' \mid\mid K2' \mid\mid K3'$$

Despite Unsplash's robust built-in search, the relevance of the images that are returned may differ. A CLIP (Contrastive Language–Image Pretrained) model is used to determine how well each image matches the user's keyword query in order to refine the results.

CLIP embeds both:

the text query Q, and each image I_i into the same latent vector space.

Text embedding

$$\mathrm{Etext} = \mathrm{fCLIP}(Q)$$

Image Embedding

$$\mathrm{Eimg}, i = \mathrm{fCLIP}(\mathrm{Ii})$$

Relevance Score

The similarity between the query and each image is computed using cosine similarity:

$$\frac{\mathrm{Etext}\ .\mathrm{Eimg,i}}{\|\mathrm{Etext}\|\|\mathrm{Eimg,i}\|} = \mathrm{Si}$$

REFERENCES

[1] A. Srivastava, A. Gupta, K. Srivastava, A. Vishwakarma, and P. Shukla, "Exploring advancements in generating images from textual descriptions using generative AI," International Journal of Scientific Research in Engineering and Management, 2024.

[2] T. Nomoto, "An updated perspective on keyword extraction methodologies," SN Computer Science, 2023.

[3] A. Bhambore, Bhagyashri, R. Pavithra, C. Tejashwini, and S. Reshma, "Text-to-image generation using generative AI," International Journal of Scientific Research in Engineering and Management (IJSREM), 2023.

[4] D. Baidoo-Anu and L. Owusu Ansah, "Education in the era of generative AI: Examining the role of ChatGPT in enhancing teaching and learning," Journal of AI, 2023.

[5] A. K. Shahade and P. V. Deshmukh, "Generative AI and its applications in natural language processing: A comprehensive review," International Journal of Computing and Digital Systems, 2023.

[6] A. B. Yadav, "An analysis on the use of image design with generative AI technologies," International Journal of Trend in Scientific Research and Development (IJTSRD), 2024.