VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

(An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering)

Department of Computer Engineering

Project Report on

Social Stories Generator: An LLM-based learning tool for specially-abled children

Submitted in partial fulfillment of the requirements of Third Year (Semester–VI), Bachelor of Engineering Degree in Computer Engineering at the University of Mumbai Academic Year 2024-25

By

Sarang Pavanaskar D12A/51

Akshat Mahajan D12A/40

Tanmay Maity D12A/42

Mohit Vaidya D12A/60

Project Mentor
Dr. Sujata Khedkar

University of Mumbai
(AY 2024-25)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

(An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering)

Department of Computer Engineering

CERTIFICATE

This is to certify that _____of Third Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on "----------" as a part of the coursework of Mini Project 2B for Semester-VI under the guidance of ----------. in the year 2024-25.

_____

Date

_____ _____

Internal Examiner External Examiner

_____	_____	_____
Project Mentor	Head of the Department	Principal
Dr. Sujata Khedkar	Dr. Mrs. Nupur Giri	Dr. J. M. Nair

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

---	---
(Signature)	(Signature)
---	---
(Sarang Pavanaskar- 51)	(Akshat Mahajan- 40)
---	---
(Signature)	(Signature)
---	---
(Tanmay Maity- 42)	(Mohit Vaidya- 60)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Dr. (Mrs.) Sujata Khedkar** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department

COURSE OUTCOMES FOR T.E MINI PROJECT 2B

Learners will be to:-

CO No.	COURSE OUTCOME
CO1	Identify problems based on societal /research needs.
CO2	Apply Knowledge and skill to solve societal problems in a group.
CO3	Develop interpersonal skills to work as a member of a group or leader.
CO4	Draw the proper inferences from available results through theoretical/ experimental/simulations.
CO5	Analyze the impact of solutions in societal and environmental context for sustainable development.
CO6	Use standard norms of engineering practices
CO7	Excel in written and oral communication.
CO8	Demonstrate capabilities of self-learning in a group, which leads to lifelong learning.
CO9	Demonstrate project management principles during project work.

ABSTRACT

The proposed system presents a novel AI-driven approach to creating personalized social stories for children with autism spectrum disorder (ASD). This system integrates five key stages: text generation, image generation, audio generation, PDF generation, and video generation. By fine-tuning large language models (LLMs) such as Gemini on a dataset of over 100 social stories, the system customizes the generated stories to match each child's specific needs. To improve social understanding and behavior of a child, the system integrates multimedia elements such as images, audio, and video. This approach demonstrates how generative AI can support education for children with special needs by automating the manual process of creating social stories and providing a personalized, multi-sensory learning experience.

Index

Title **Page no.**

Chapter 5: Implementation of the Proposed System

5.1. Methodology Employed

5.2 Algorithms and flowcharts

5.3 Dataset Description

Chapter 6: Testing of the Proposed System

6.1. Introduction to testing

6.2. Types of tests Considered

6.3 Various test case scenarios considered

6.4. Inference drawn from the test cases

Chapter 7: Results and Discussion

7.1. Screenshots of User Interface (GUI)

7.2. Performance Evaluation measures

7.3. Input Parameters / Features considered

7.4. Graphical and statistical output

7.5. Comparison of results with existing systems

7.6. Inference drawn

Chapter 8: Conclusion

8.1 Limitations

8.2 Conclusion

8.3 Future Scope

References

Appendix

1. Research Paper Details

a. List of Figures
b. List of Tables
c. Paper Publications
d. Certificate of publication
e. Plagiarism report
f. Project review sheets

2. Competition certificates from the Industry (if any)

Appendix

a. List of Figures

Figure Number	Heading	Page no.

b. List of tables

Table Number	Heading	Page no.

c. Paper Publications :-

1. **Draft of the paper published.**
2. **Plagiarism report of the paper published /draft**
3. **Certificate of the paper publication**
4. **Xerox of project review sheets**

Chapter 1: Introduction

1.1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects a child's ability to communicate, socialize, and adapt to everyday situations. Children diagnosed with ASD often face challenges in understanding social cues, expressing emotions, managing transitions, and engaging in routine activities. To support them in navigating these challenges, therapists and special educators frequently use a tool known as *Social Stories*. These are short, structured narratives designed to explain specific social situations, behaviors, or expectations in a simple and relatable manner. Traditionally, these social stories are created manually for each child by educators, psychologists, or parents, often incorporating text and images tailored to a specific situation—such as going to school, brushing teeth, or sharing toys.

While social stories have proven to be effective in helping children with ASD understand and respond to social scenarios, the manual process of creating these stories can be time-consuming, repetitive, and not always feasible, especially when there's a need for regular updates, multilingual delivery, or personalized content. Moreover, the lack of engaging visual or auditory elements can limit the impact of these stories on certain children who respond better to multi-sensory learning. In today's world, where artificial intelligence (AI) and machine learning (ML) have made personalization and automation more accessible, there is an excellent opportunity to enhance and scale this process using generative AI technologies.

This project presents an AI-driven system for generating personalized social stories for children with ASD. The system is designed to work in five stages: text generation, image generation, audio narration, PDF compilation, and video creation. Each stage is powered by cutting-edge generative AI models and tools, enabling the system to produce high-quality, consistent, and engaging content tailored to each child's developmental needs. The goal is to transform the way social stories are created—making them more accessible, dynamic, and effective in supporting children's social development.

1.2 Motivation

The inspiration behind this project stems from the significant gap that exists in creating scalable and personalized educational content for children with special needs, particularly those on the autism spectrum. In many schools, therapy centers, and homes, caregivers and educators struggle with the constant demand to prepare new social stories for different children, each with unique behavioral goals, language preferences, and sensory sensitivities. This process is not only repetitive but also resource-intensive, as it requires dedicated time, creativity, and technical know-how to produce content that is both visually appealing and behaviorally relevant.

Moreover, while technology has transformed mainstream education through digital learning platforms and AI tutors, the same level of innovation has not fully reached the realm of special education. Most of the existing tools lack the intelligence to adapt content to individual needs, and even fewer offer support across multiple media formats. Children with ASD often benefit from multi-sensory learning—combining visual, auditory, and textual information—and are more engaged when the content reflects their interests or routines. This motivates the need for a system that not only generates social stories but does so intelligently, interactively, and inclusively.

Our motivation is to harness the power of AI—particularly large language models (LLMs) like Gemini and image/audio generation tools—to simplify the process of creating high-quality social stories. This approach reduces the manual burden on educators and provides children with personalized, engaging stories that can better support their social and behavioral learning.

1.3 Problem Definition

The core problem addressed by this project is the **lack of a scalable, intelligent, and multimedia-supported system** for generating personalized social stories for children with Autism Spectrum Disorder. Traditional methods rely heavily on manual content creation, which is time-consuming, limited in adaptability, and often lacks consistency in quality and visual presentation. Furthermore, the static nature of printed stories fails to engage many children who respond better to dynamic, animated, or interactive content.

The proposed system seeks to automate the generation of social stories using generative AI models capable of understanding the learning objective and creating suitable narratives. It then visualizes these stories with context-aware illustrations, converts them into speech

using text-to-speech engines, compiles them into printable PDFs, and finally generates animated videos with synchronized narration. Additionally, the system includes multilingual support and quiz integration to reinforce learning.

The aim is to create a comprehensive pipeline that transforms a basic input (like a behavioral goal or activity) into a full-fledged, multi-sensory story experience, customized for a child's needs. This system addresses the current inefficiencies in social story creation and opens up new possibilities for inclusive, adaptive education.

1.4 Existing Systems

There are several existing tools and software that support the creation of social stories or visual aids for children with special needs. For instance, platforms like **Boardmaker®** and **Widgit** provide libraries of pre-designed symbols and templates that educators can use to construct visual stories. These tools are widely used in special education classrooms but are largely template-based and require manual input for text and visuals, offering limited scope for automation or personalization.

Mobile applications like **Social Story Creator & Library** or **Pictello** allow parents and teachers to build custom stories using images and recorded audio. While these apps improve accessibility and usability, they still depend on manual effort and do not leverage AI to generate content dynamically. Moreover, maintaining consistency in characters or styles across different stories is difficult without a unified generative framework.

Some tools offer basic translation or audio support, but they are not integrated into a single pipeline. Also, these systems generally do not offer video generation, character consistency, or real-time personalization based on user preferences or developmental goals. As a result, the potential of using AI and automation in this space remains largely untapped.

1.5 Lacuna of the Existing Systems

Despite their usefulness, the current solutions for social story creation fall short in several critical areas. Firstly, **manual content creation** remains a major bottleneck. Teachers or parents must write, edit, and illustrate each story, which is impractical when working with multiple children or when stories need to be updated frequently. Secondly, **personalization is extremely limited**—existing tools do not adapt to a child's individual learning needs, preferences, or challenges.

Another major shortcoming is the **lack of consistency in visuals**. When using clipart or third-party images, it's difficult to maintain a uniform look for a character across different pages or scenarios. Additionally, these systems generally lack **multilingual support**, making it hard for non-English-speaking children and parents to benefit from them. Furthermore, **audio and video integration is either missing or minimal**, depriving children of a more immersive learning experience. There are also no built-in quiz mechanisms to check understanding or reinforce behavioral concepts.

All these gaps highlight the need for a more intelligent, fully integrated system that automates the story creation process while providing visually consistent, personalized, and multimodal learning outputs.

1.6 Relevance of the Project

This project is highly relevant in the current educational and technological landscape, particularly with the increasing prevalence of ASD among children. According to global statistics, the number of ASD diagnoses has been steadily rising, creating a growing demand for effective and scalable educational interventions. At the same time, advancements in AI and machine learning have made it possible to generate human-like text, create realistic images, produce natural-sounding speech, and even animate entire videos—all in an automated manner.

By bringing these capabilities together, our system addresses the urgent need for accessible and personalized learning tools in special education. It not only reduces the workload of educators and therapists but also ensures that every child receives content tailored to their unique developmental journey. Moreover, the project aligns with the growing trend of **inclusive and tech-enabled education**, making learning more equitable for neurodivergent children.

With features such as multilingual support, audio narration, interactive quizzes, and printable/exportable formats, this system empowers educators, therapists, and parents to engage children in a more meaningful and effective way. It also has potential applications beyond ASD—supporting children with learning disabilities, speech delays, or behavioral challenges. In essence, the project leverages the latest in AI to offer a solution that is not only technologically innovative but socially impactful as well.

Chapter 2: Literature Survey

A. Overview of Literature Survey

The field of generative AI has witnessed rapid advancements, especially in domains like text generation, image synthesis, and multimodal applications. These developments have opened avenues for solving challenges in personalized education, including for children with Autism Spectrum Disorder (ASD). Social stories—narratives that help children with ASD understand and navigate social situations—have traditionally been developed manually by caregivers or psychologists. However, manual story creation is time-consuming and lacks scalability. This literature survey aims to explore existing systems, methodologies, and tools relevant to AI-driven story generation, text-to-image synthesis, and multimodal learning platforms. Through this, we identify the strengths and limitations of related works, outline current trends, and position our project within the research gap.

B. Related Works

The following section reviews and analyzes significant research works that focus on LLMs, generative AI applications, text-to-image synthesis, empathy modeling, and vision-language understanding. These papers provide the theoretical foundation and motivation for our AI-based social story generator that uses text, images, video, audio, and quizzes tailored for children with ASD.

2.1 Research Papers Referred

1. Sarid: Arabic Storyteller Using a Fine-Tuned LLM and Text-to-Image Generation (2024)

a. Abstract of the Research Paper:

This research developed an AI-based Arabic story generator tailored for children. It fine-tuned OpenAI's Davinci language model on 527 Arabic children's stories and incorporated Midjourney for text-to-image generation. The tool accepts user inputs such as the character, tone, and theme to dynamically create personalized stories along with visuals. It emphasizes maintaining character consistency throughout the narrative using prompt engineering and careful image synthesis.

b. Inference Drawn:

The study showcases the practical application of fine-tuned LLMs combined with image generation tools. It highlights the importance of user input in achieving personalization and demonstrates the benefits of integrating text and images. However, it also exposes challenges in character consistency and reliance on prompt quality. These insights influenced our design to ensure better consistency across media and deeper personalization for children with ASD.

2. A Survey of Generative AI Applications (2024)

a. Abstract of the Research Paper:

This paper systematically categorizes over 350 generative AI applications, classifying them into domains like text, image, video, and audio. It also identifies the technological stacks (GANs, VAEs, LLMs, Diffusion models) used in building these applications. The study highlights key industry tools and academic frameworks driving the generative AI revolution.

b. Inference Drawn:

This survey provided a broader landscape of where generative AI is heading and served as a benchmark for identifying the most stable and scalable technologies for our use case. It validated our decision to use LLMs for story generation and multimodal integration. The wide scope of the paper also hinted at potential integrations we can explore in the future, like gamified learning and AI-driven tutors.

3. Deep Learning Methodology Converts Text to Image (2022)

a. Abstract of the Research Paper:

The paper explores how GANs can be utilized to synthesize images from textual inputs. Using TensorFlow and NLTK, a model was trained to generate images from descriptions. It applied the GAN-CLS algorithm and was evaluated using the Oxford-102 flower dataset. The model used a GUI built with PySimpleGUI for user interaction.

b. Inference Drawn:

Though the dataset and theme were unrelated to social stories, the methodology behind text-to-image synthesis directly supported our image generation module. The study also reinforced the necessity of having a robust textual representation before triggering the image generation process. The challenges it exposed—such as semantic misalignment between text and image—led us to ensure tighter control between our text generation and image synthesis pipelines.

4. LLM-DetectAIve: A Tool for Fine-Grained Machine-Generated Text Detection (2024)

a. Abstract of the Research Paper:

This tool classifies text into four categories: human-written, machine-generated, machine-humanized, and human-polished. It uses models like RoBERTa and DeBERTa for classification and introduces a new dataset combining human and machine-generated texts.

b. Inference Drawn:

While this tool is not focused on story generation, it demonstrates the granularity possible in text classification using LLMs. This inspired us to look into how fine-grained classification techniques can help assess the emotional quality and simplicity of the generated stories, especially for children with varying cognitive abilities.

5. HEART-felt Narratives: Tracing Empathy and Narrative Style in Personal Stories with LLMs (2024)

a. Abstract of the Research Paper:

This study examined how different narrative styles affect empathy levels in storytelling. It introduced the HEART taxonomy and analyzed over 2,600 crowd-sourced responses using LLMs to assess the presence of empathy-related elements.

b. Inference Drawn:

This research is highly relevant to our goal of making stories emotionally resonant for children with ASD. The idea of analyzing narrative styles using AI encouraged us to experiment with tone, empathy, and personalization when generating stories. It informed our story formatting engine, especially when generating situations like making friends or visiting the doctor.

6. How Well Can Vision-Language Models See Image Details? (2024)

a. Abstract of the Research Paper:

The paper evaluates how well Vision-Language Models (VLMs) understand pixel-level details in images. It introduces a new pixel-value prediction (PVP) task and fine-tunes VLMs accordingly, demonstrating improvements in image segmentation and visual perception.

b. Inference Drawn:

This research validated our concern regarding the fidelity of image details. For social stories, consistent visual storytelling is crucial. It led us to favor image models that offer pixel-level customization and clarity, especially when a child needs to recognize expressions, environments, or objects.

7. Muse: Text-To-Image Generation via Masked Generative Transformers (2023)

a. Abstract of the Research Paper:

Muse is a text-to-image model that uses masked modeling on image tokens for improved decoding efficiency. It outperforms diffusion and autoregressive models on standard benchmarks and supports advanced editing without re-training.

b. Inference Drawn:

Muse introduced the concept of fine-grained image editing without retraining, which aligns with our requirement to let users regenerate individual scenes without disrupting the whole story. This gave us ideas for future versions of the system to enable story section editing.

2.2 Patent Search

Though limited, existing patents primarily focus on static storybook generation tools or apps for children with speech disorders. Some patent applications involve educational content creation using templates. However, none offer fully automated, AI-driven, multimodal social story generation tailored to ASD children with real-time feedback and cultural adaptability. This highlights a potential IP opportunity for our system.

2.3 Inference Drawn

From both research literature and the limited scope of patents, several important takeaways emerged. First, the integration of LLMs and generative models is an emerging field in education, with limited application to neurodiverse learners. Second, there is an unmet need for tools that provide multilingual, personalized, and media-rich content. Third, combining real-time interaction, user feedback, and cross-format story coherence remains a research and development gap.

2.4 Comparison with the Existing System

Aspect	Existing Systems	Our Proposed System
Story Customization	Limited user input, mostly template-based	Personalized by child's age, comprehension, and needs
Multimedia Integration	Mostly text-based or image-only	Full multimedia: text, image, audio, video, quizzes
Cultural and Language Support	Primarily English, lacks diversity	Supports multilingual generation and translation
Feedback and Adaptability	No real-time feedback loops	Planned real-time feedback from parents/educators
Automation	Manual or semi-automated content creation	Fully automated story and media generation
Consistency Across Formats	Inconsistent image-text pairings	Same theme, tone, and characters maintained throughout
Dataset	No publicly curated social story datasets	Fine-tuned on 100+ curated stories
Usability and Scalability	Often non-intuitive and unscalable	Designed on Streamlit, highly usable and scalable

Chapter 3: Requirement Gathering for the Proposed System

3.1 Introduction to Requirement Gathering

Requirement gathering is a foundational step in any software development life cycle (SDLC) and is crucial for aligning the system's functionality with end-user needs. In the context of our AI-driven personalized social story generation system for children with Autism Spectrum Disorder (ASD), this phase focused on identifying the specific requirements of educators, therapists, caregivers, and children. By understanding the unique challenges faced in teaching social behaviors to children with ASD, we designed a system that automates and personalizes the creation of multimedia-rich social stories.

Our project is implemented entirely using **Streamlit**, a lightweight Python-based web framework that allows rapid development of a user-friendly interface. The backend logic is also built in **Python**, leveraging multiple machine learning APIs and media processing libraries. **Firebase** is used for secure user authentication, and **Large Language Model (LLM) APIs** like Gemini or GPT have been integrated for story text generation. This architecture supports an end-to-end pipeline for generating stories with text, image, audio, and video, customized to each child's context and needs.

3.2 Functional Requirements

Functional requirements specify the core capabilities of the system—the tasks it must perform to fulfill its purpose. The key functional requirements of our system include:

- **Text Generation:** The system should allow the user to provide a topic or target behavior. Using an LLM API, the system should generate a structured, child-friendly social story that aligns with the needs of a child with ASD.
- **Image Generation:** For each part of the story, the system should use generative models to create relevant and consistent comic-style images. These images are generated via API calls and are contextually mapped to story paragraphs.
- **Audio Generation:** The system converts the generated text into speech using a Text-to-Speech (TTS) engine, supporting multilingual narration for better accessibility.

- **PDF Generation:** The entire story, including images and text, is compiled into a downloadable PDF format that parents or educators can print or store.
- **Video Generation:** A video version of the story is generated using Python libraries like **MoviePy**, combining the audio and images in a timeline for a complete audiovisual experience.
- **Quiz Module:** After story presentation, the system generates quizzes (true/false, MCQs) to reinforce understanding and assess learning outcomes.
- **User Authentication:** Firebase Authentication is used to securely manage user accounts and sessions via email/password or phone-based OTP.
- **Interactive UI:** Streamlit provides the user interface for all interactions, including input forms, previews of generated content, and download buttons for PDFs or videos.

3.3 Non-Functional Requirements

Non-functional requirements ensure the system performs well under various conditions and constraints. For our application, the most important non-functional aspects are:

- **Responsiveness and Speed:** Since LLMs and image generation APIs can take a few seconds to return outputs, the UI is designed to show progress indicators and keep users engaged during waits.
- **Usability:** Streamlit's component-based UI is intuitive and requires no technical expertise. Story previews, quizzes, and downloads are all available with a few clicks.
- **Maintainability:** The entire codebase is modularized in Python, making it easy to update APIs or switch to new LLM providers or TTS tools without rewriting the entire pipeline.
- **Security:** Firebase Authentication ensures secure logins, and user-uploaded data (if any) is stored securely using Firebase Storage with access rules.
- **Multilingual Capability:** Text, audio, and quiz generation all support multiple languages to cater to diverse user groups across geographies.
- **Accessibility:** The use of audio narration, visual images, and simplified text ensures accessibility for children with various learning needs.

3.4 Hardware, Software, Technology and Tools Utilized:

Hardware Requirements:

Since Streamlit apps are lightweight and browser-based, hardware requirements are minimal:

- **Development Environment:**
  - Processor: Intel i5 or equivalent
  - RAM: 8GB minimum
  - GPU: Optional but recommended for faster media processing (if running video/image generation locally)

- **Deployment Server (Optional):**
  - For cloud deployment, platforms like Google Cloud, Streamlit Community Cloud, or AWS EC2 can be used with basic Python support.

Software and Tools:

- **Frontend & Backend:**
  - **Streamlit** (Python-based web UI framework)
  - **Python** (Core programming language for backend, ML integration, and media generation)

- **Authentication & Storage:**
  - **Firebase Authentication** (Email/password, phone OTP login)
  - **Firebase Storage/Database** (For storing media, user settings, or story logs)

- **Utilities:**
  - **Git & GitHub** (Version control)
  - **LangChain** (For chaining LLM responses, if used)
  - **NumPy, PIL, OpenCV, base64** (for image and video preprocessing in Python)

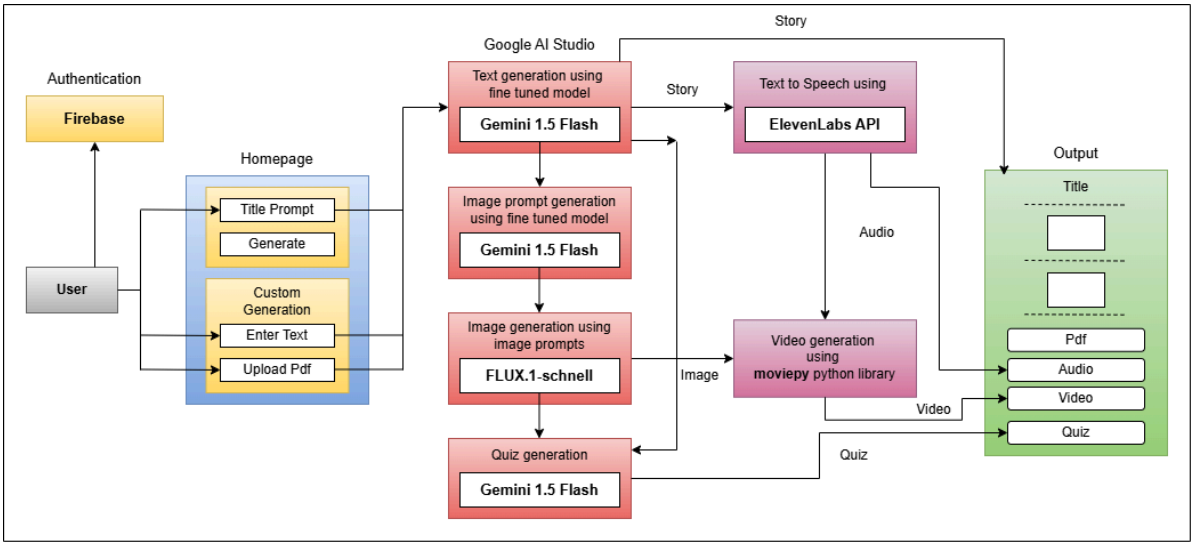
3.5 Constraints

Though powerful, the system has a few constraints that must be acknowledged:

- **Internet Dependency:** Since the LLM, TTS, and image APIs are cloud-based, the system requires a stable internet connection for story generation to function.
- **API Limits & Costs:** Third-party APIs often have rate limits or pricing plans that can restrict free usage, especially in high-demand settings like schools.
- **Model Latency:** Generating multiple images or longer stories may take time, and ensuring real-time performance may require asynchronous processing or caching.
- **Character Consistency:** Image generation tools may struggle with keeping the same character consistent across scenes, which can affect the story's coherence.
- **Language Support Variability:** While English support is excellent, some Indian languages may have limited or lower-quality support in current TTS or image APIs.
- **Limited Offline Support:** As Streamlit is web-based, offline usage is limited unless users download the PDFs or videos for use without internet access.

Chapter 4: Proposed Design

4.1 Block diagram of the system



4.2 Modular Design of the System

- **Authentication Module**: Handles user verification via Firebase.
- **Input Handling Module**: Accepts title, custom text, or PDF input.
- **Text Generation Module**: Uses Gemini 1.5 Flash to create social stories.
- **Image Generation Module**: Converts prompts into visual illustrations.
- **Audio Module**: Converts story text into speech using ElevenLabs.
- **Video Module**: Combines images and text to create story videos.
- **Quiz Generation Module**: Auto-generates questions based on content.
- **Output Management Module**: Displays or downloads the final product.

Each module performs a unique function but communicates with others through well-defined interfaces, promoting scalability and ease of debugging.

4.3 Detailed Design

- **Text Generation Module**:
    - Input: Prompt
    - Process: Passed to Gemini 1.5 Flash via Google AI Studio
    - Output: Full social story text

- **Image Module**:
    - Uses image prompts from Gemini model
    - Calls Flux DreamBooth on Hugging Face
    - Supports multithreading for speed

- **Audio & Video Module**:
    - Uses ElevenLabs API to get speech in MP3
    - Uses `moviepy` to merge images + audio into MP4

Chapter 5: Implementation of the Proposed System

5.1 Methodology Employed

The proposed system integrates multiple AI components into a cohesive pipeline designed to generate personalized social stories with multimedia enhancements for children with special needs. The implementation employs a combination of Large Language Models (LLMs), diffusion-based image generation, and multimedia integration to produce an engaging and accessible learning tool.

The methodology includes:

1. **User Authentication and Input**:
   Authentication is handled via **Firebase**, ensuring secure access. Users can generate stories through title-based prompts, custom text input, or by uploading a PDF.
2. **Story and Prompt Generation**:
   Text generation and image prompt creation are handled using **Gemini 1.5 Flash** via **Google AI Studio**, tailored with fine-tuned models to suit the context of social learning.
3. **Image Generation**:
   The generated prompts are passed to a diffusion model, **F1L1X-kandinsky-3.0**, to create relevant, child-friendly illustrations.
4. **Audio Generation**:
   Generated stories are converted into multilingual audio using the **ElevenLabs API**, enhancing accessibility for children with different learning needs.
5. **Quiz Creation**:
   Comprehension-based quizzes are automatically generated using **Gemini 1.5 Flash**, based on the story content.
6. **Video Generation**:
   Images, narration, and story text are integrated into a video format using the **moviepy** Python library.
7. **Final Output Compilation**:
   All elements — the story text, illustrations, audio narration, video, and quiz — are compiled into downloadable formats: **PDF**, **Audio**, **Video**, and **Quiz Module**.

5.2 Algorithms and Flowcharts

The upgraded system integrates additional functionality for enhanced educational engagement, including video generation and quiz creation. The system follows a modular, multistage workflow that leverages LLMs, TTS, and media generation libraries to produce a comprehensive output suite — text, audio, images, video, and quizzes.

The user initiates interaction by choosing from various input methods — title-based prompt, custom text input, or PDF upload. After Firebase-based authentication, the system processes the input through multiple fine-tuned models to generate diverse outputs.

5.3 Dataset Description

The effectiveness of the StoryGPT system hinges on the availability of a high-quality dataset tailored for fine-tuning the Gemini large language model. The dataset was specifically curated to enable the generation of contextually appropriate and personalized social stories for children, particularly those with Autism Spectrum Disorder (ASD).

5.3.1 Data Collection

The dataset was compiled through extensive web scraping from publicly accessible educational platforms and repositories containing social stories. These sources included a broad array of behavioral scenarios and life-skills content, focusing on topics such as:

- Communication and interpersonal skills
- Emotional self-regulation
- Personal safety and hygiene
- Appropriate behavior in social environments

This diversity ensured that the model could be trained on realistic, relatable situations that children may encounter in day-to-day life.

5.3.2 Data Format and Preprocessing

Each story entry in the dataset followed a structured format, consisting of:

- **Title** – A concise descriptor of the social scenario (e.g., *"Going to School"*, *"Making New Friends"*).
- **Story** – A narrative segment offering guidance on how to behave appropriately in the given situation.

The collected data was preprocessed to remove inconsistencies, standardize language, and align with the Gemini model's fine-tuning requirements. The final dataset was structured as a table with two columns:

Prompt (Title)	Target Output (Story)
Making New Friends	A story explaining how to introduce oneself politely...
Going to the Doctor	A story that describes what happens during a check-up...

This format ensured compatibility with the input-output paradigm required for supervised learning during the fine-tuning process.

5.3.3 Ethical Considerations

No personal or sensitive user data was used during dataset creation. All content was either publicly available or synthetically generated to maintain ethical integrity and respect privacy standards.

Chapter 6: Testing of the Proposed System

6.1 Introduction to Testing

Testing is a vital stage in software development that ensures the system functions correctly, meets user requirements, and operates reliably under various conditions. In the proposed system—an AI-driven social story generator for specially-abled children—testing was conducted to validate the accuracy of story generation, correctness of multimedia integration, and user interface functionality. Since the system incorporates multiple components like text, image, audio, and video generation, each module was rigorously tested to ensure a smooth, personalized, and immersive learning experience.

6.2 Types of Tests Considered

Several testing techniques were used to evaluate both the backend and frontend components of the system:

- Unit Testing: Ensured that each module (e.g., story generation, audio synthesis, video generation) functioned independently without errors.
- Integration Testing: Verified the end-to-end workflow from input prompt to the final multimedia output, ensuring smooth communication between modules.
- Functional Testing: Validated that each feature (e.g., text generation, quiz creation, audio playback) works as intended for various input types (text, title, or PDF).
- Performance Testing: Measured system responsiveness, particularly the time taken to generate each content format.
- Usability Testing: Evaluated by educators and therapists to ensure the interface was intuitive and accessible for non-technical users.
- Cross-Platform Testing: Tested on multiple browsers and devices to ensure compatibility and responsiveness.

6.3 Various Test Case Scenarios Considered

The following representative test scenarios were designed and executed:

Test Case	Input	Expected Output	Status
TC01	Story title: "Brushing Teeth"	Generated story text relevant to hygiene with appropriate images and narration	✅ Pass
TC02	Uploaded PDF with story	Extracted text, generated images, audio and video correctly	✅ Pass
TC03	Input language: Hindi	Complete output generated in Hindi (text + audio)	✅ Pass
TC04	Quiz Generation	Multiple relevant quiz questions based on generated story	✅ Pass
TC05	Video Generation	Story images combined with synced narration in MP4	✅ Pass
TC06	Audio Pronunciation	Multilingual narration was clear and correctly pronounced	✅ Pass
TC07	Simultaneous Access	Multiple users generated stories concurrently without lag	✅ Pass
TC08	Missing Input	System prompts user and handles errors gracefully	✅ Pass

6.4 Inference Drawn from the Test Cases

Testing revealed that the system is robust, modular, and capable of handling diverse inputs effectively. All core features—text generation, image creation, audio synthesis, PDF compilation, video output, and quiz generation—were validated successfully. Usability feedback confirmed that even non-technical users found the system intuitive and efficient. The system's average processing time (under 60 seconds for full content generation) met real-world usability expectations.

While minor limitations such as API latency and image character inconsistency were observed, they did not hinder the core functionality. The successful execution of test cases supports the conclusion that the system is production-ready for educational environments, offering a scalable solution for personalized learning among neurodiverse children.

Chapter 7: Results and Discussion

7.1 Screenshots of User Interface (GUI)

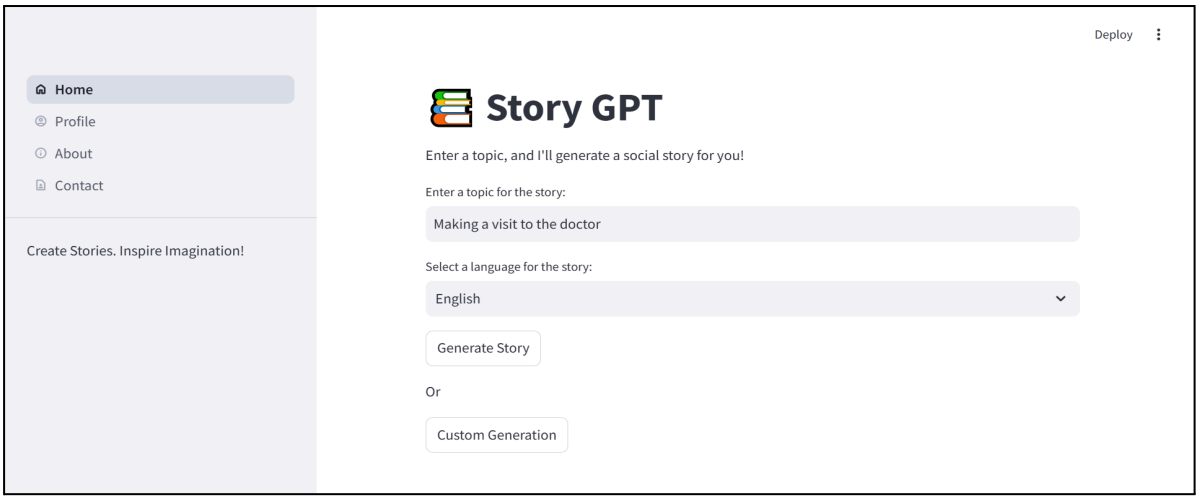
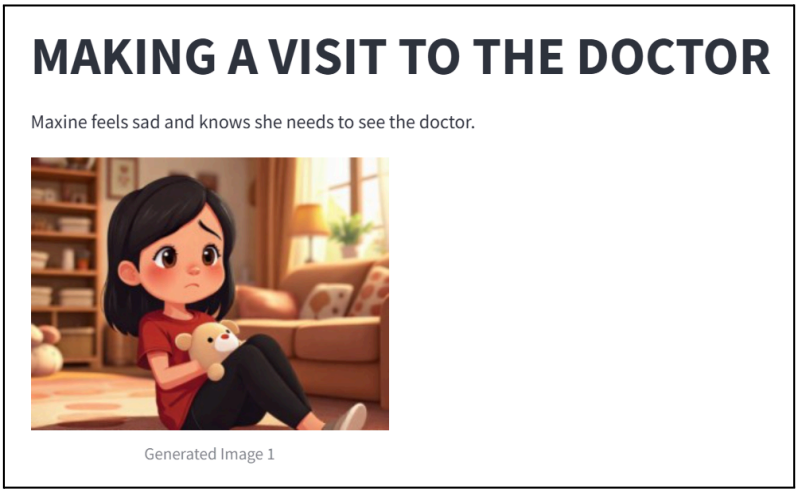
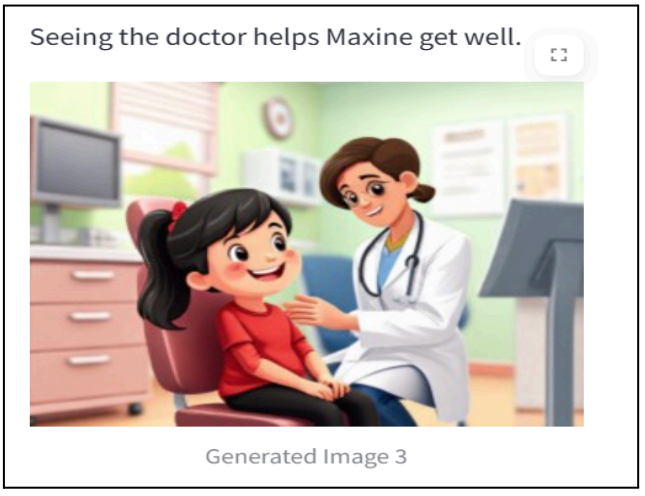


fig: homepage of social story generator





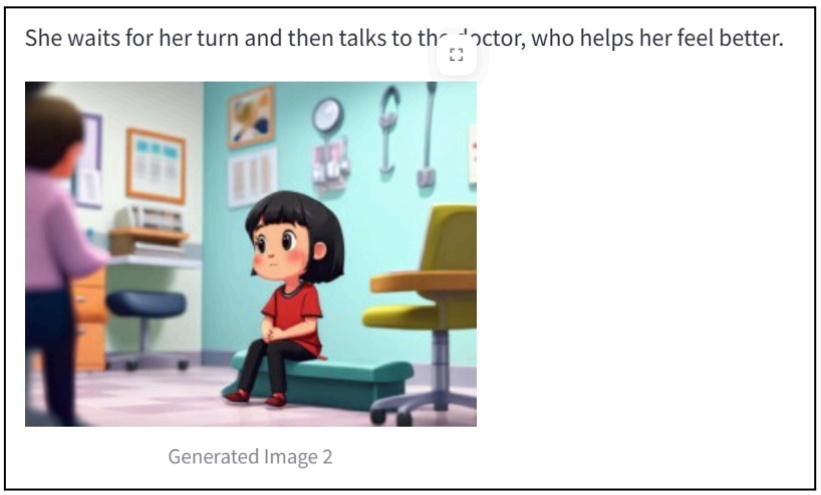


fig: story text and image generation

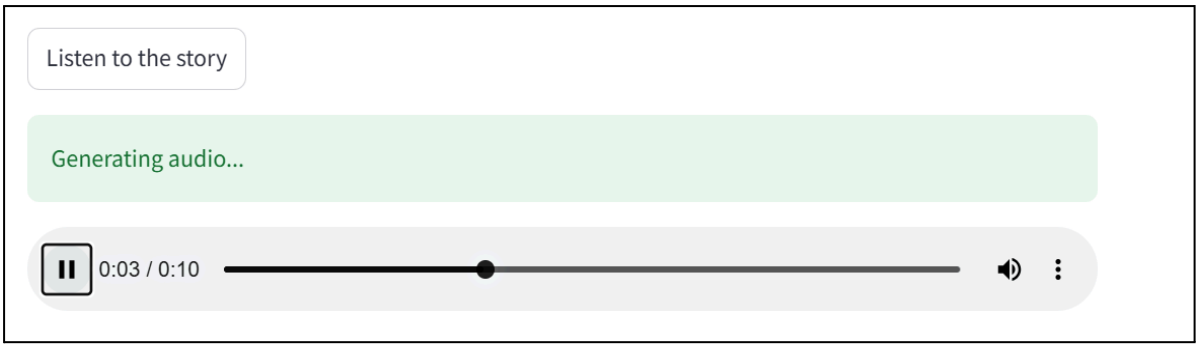


fig: audio output of story

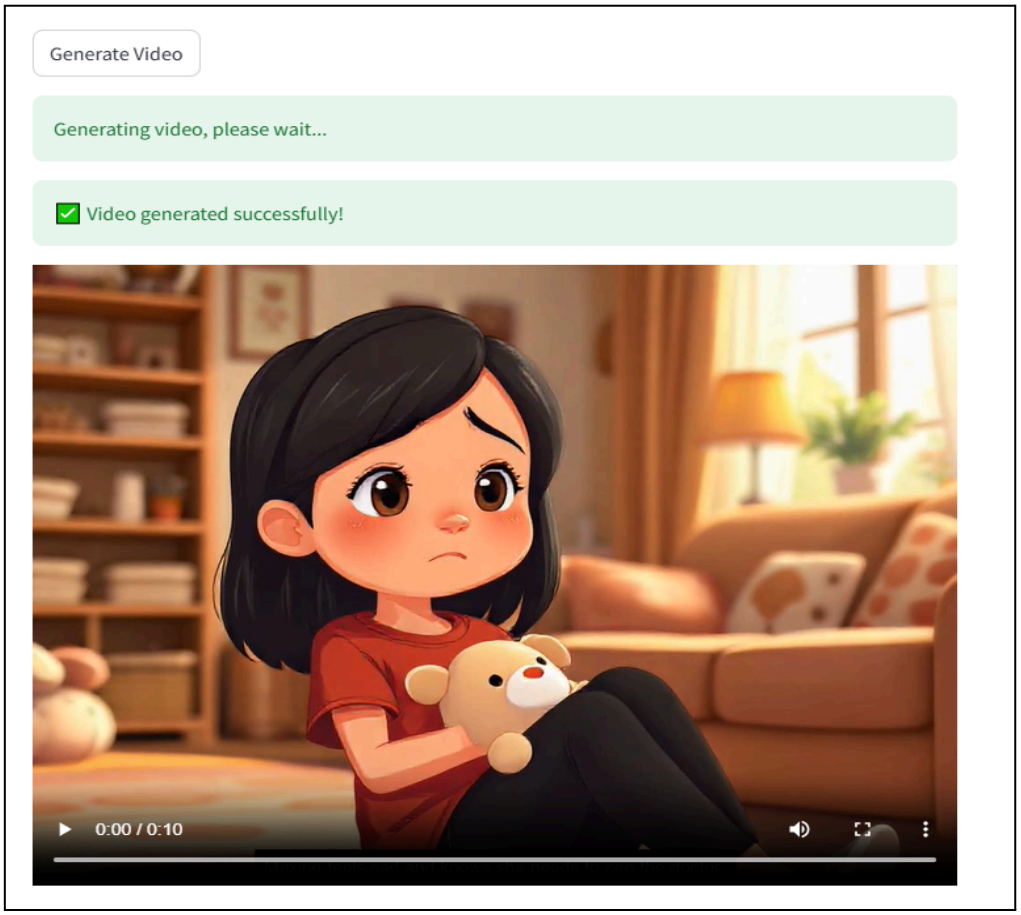


fig: video output of story

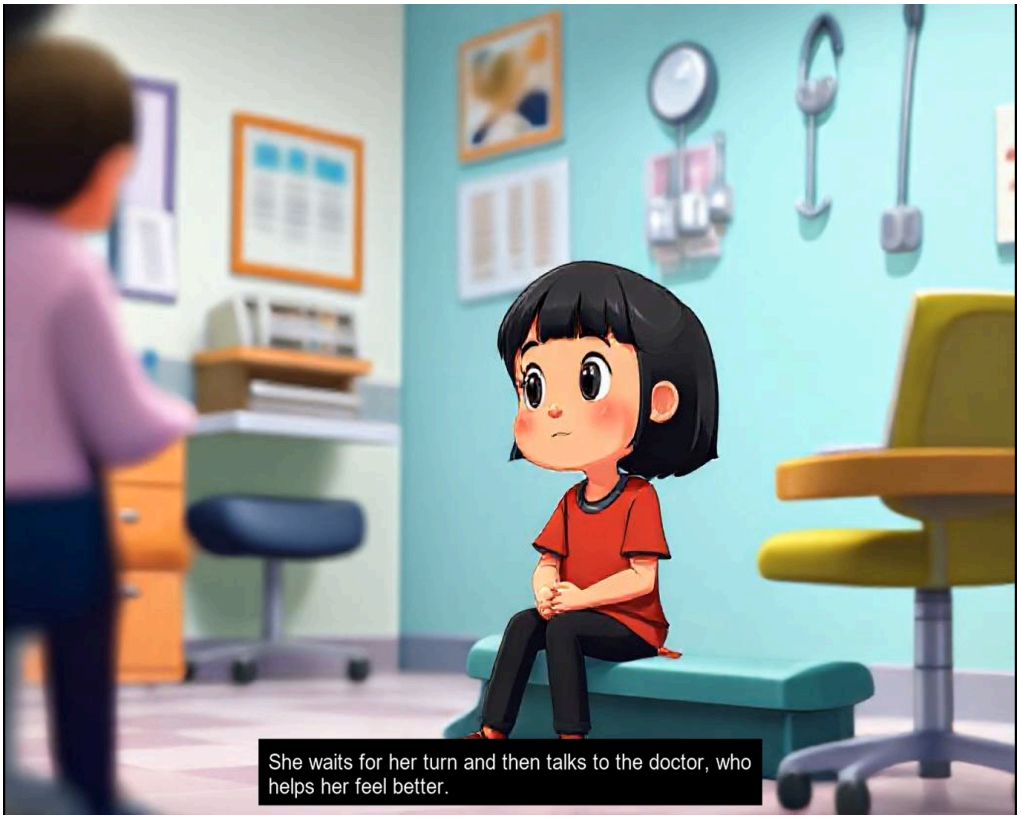
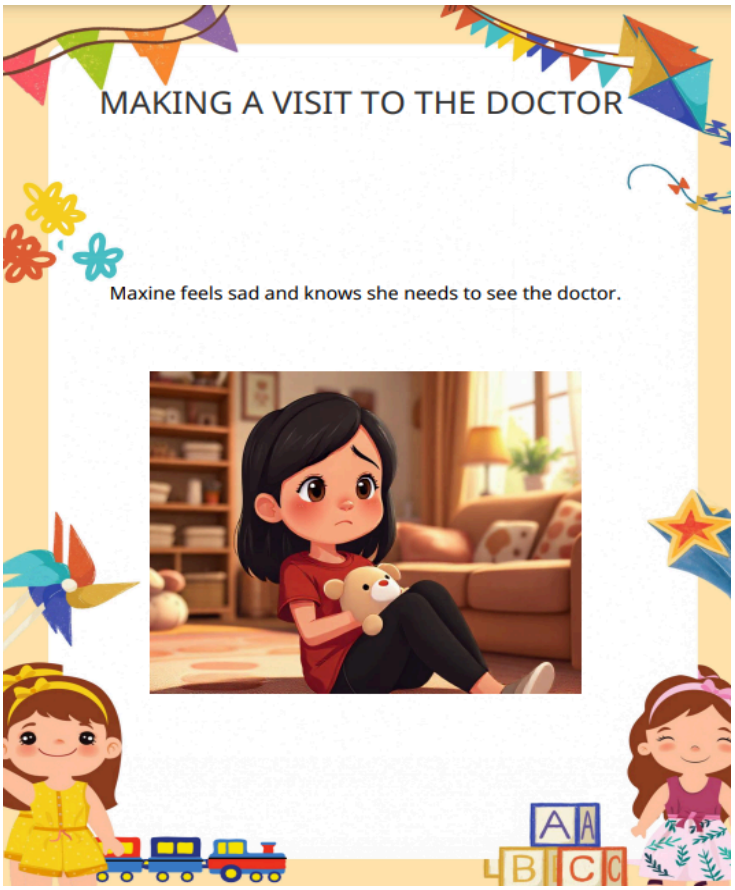


fig: video playing with subtitles



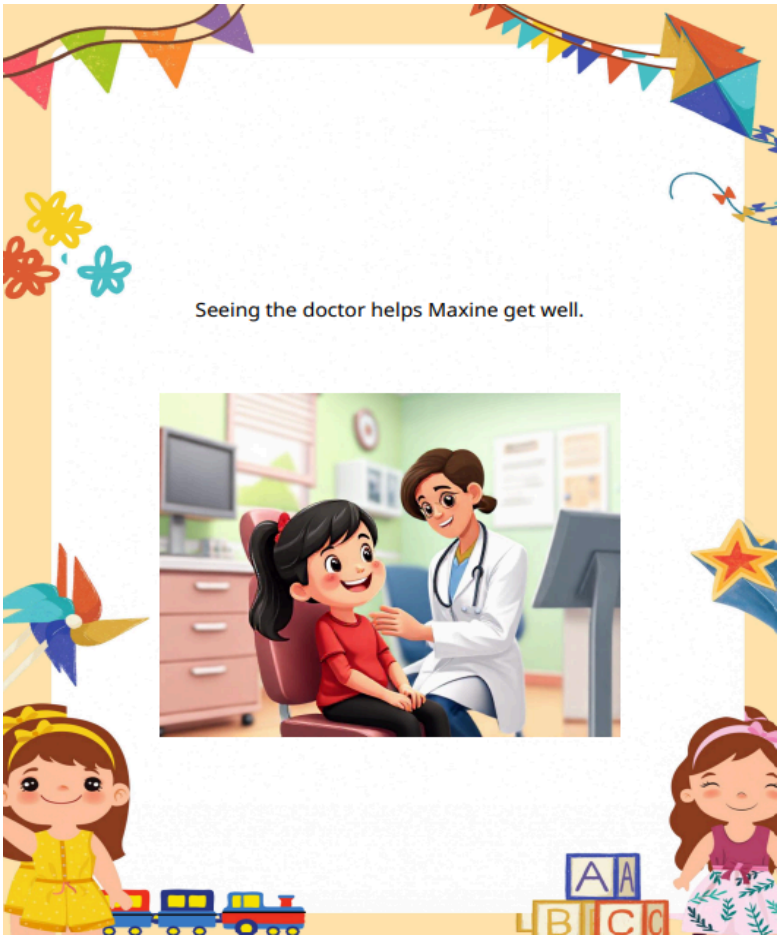
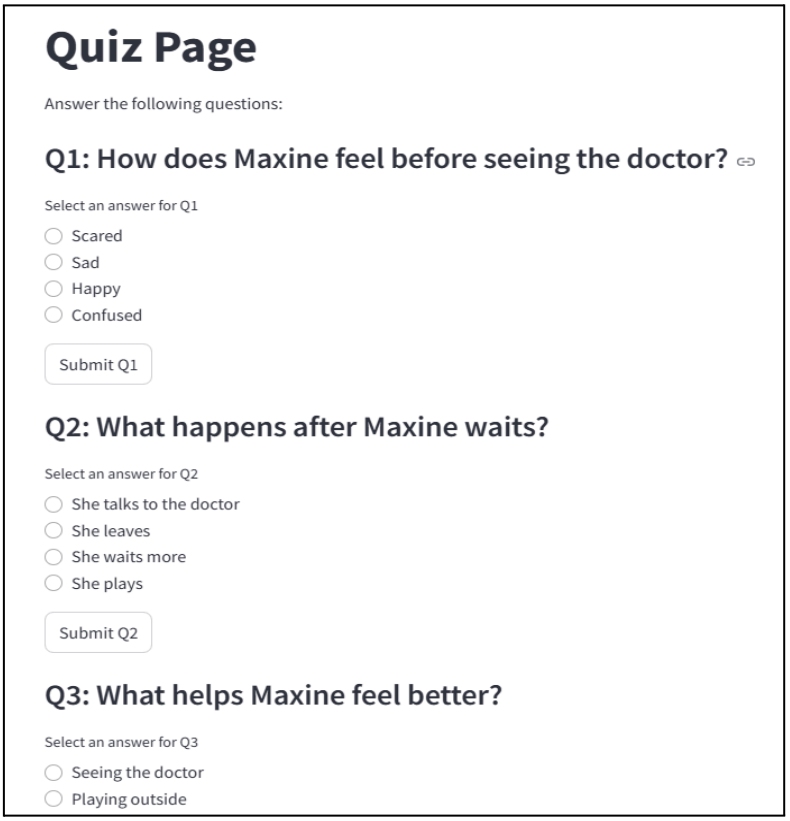


fig: pdf output of social story



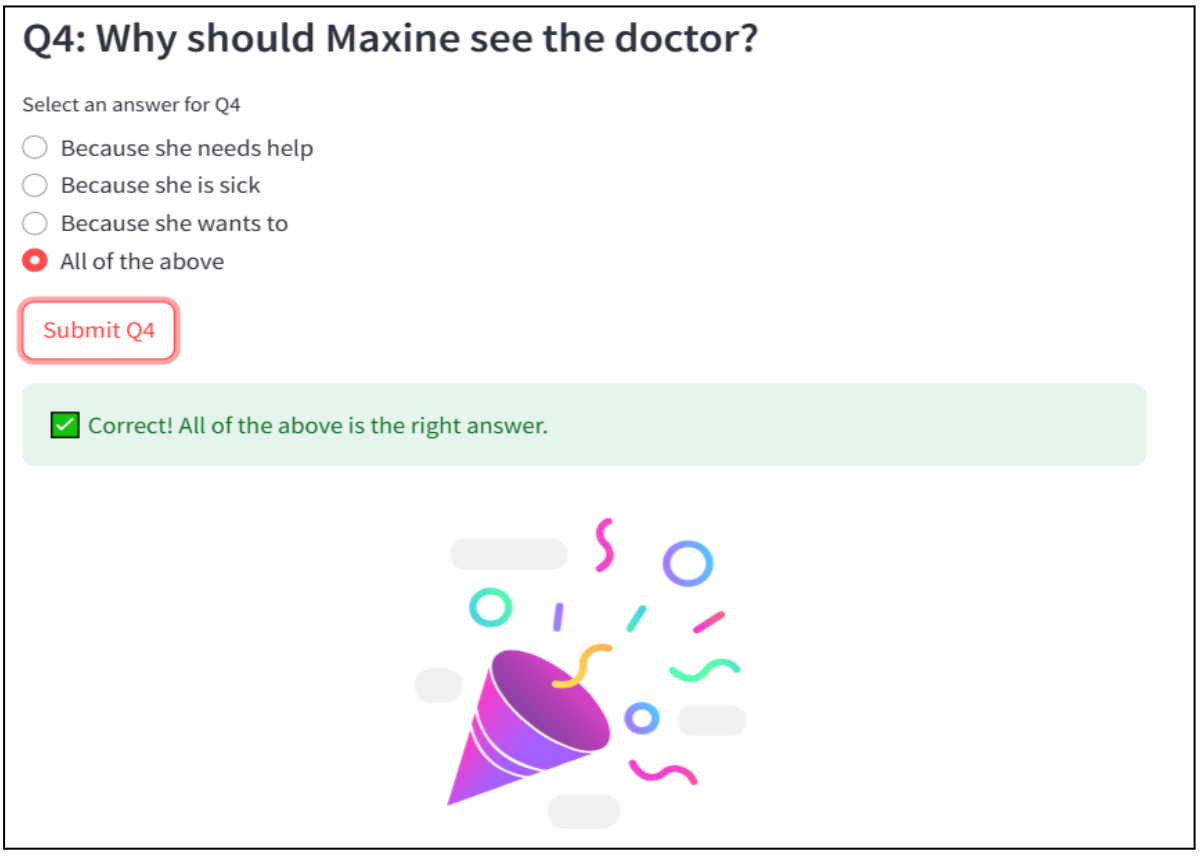


fig: quiz assessment based on generated social story

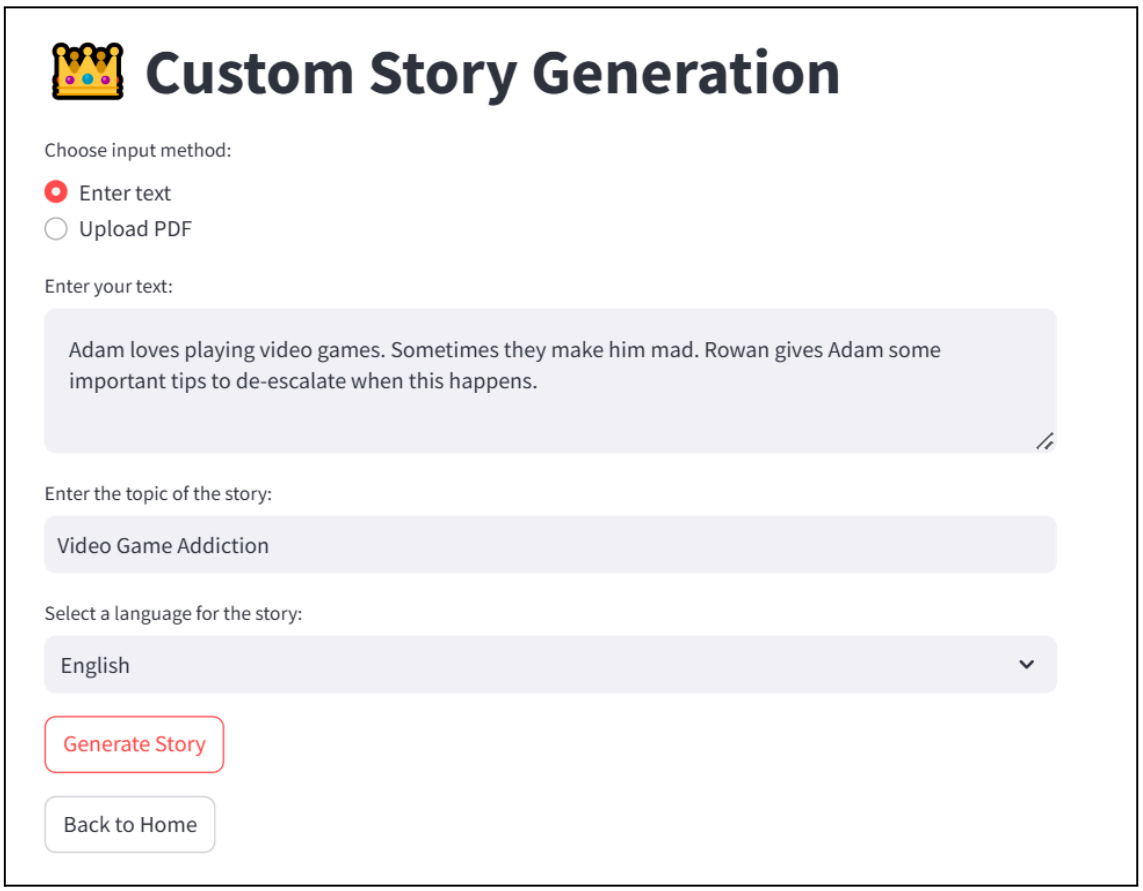


fig: Customized Social Story Generation by entering text or uploading any pdf

fig: Example of Multilingual output of social story

7.2 Performance Evaluation Measures

To assess the quality and effectiveness of the system, various performance evaluation metrics were employed. The most important metric was **text relevance**, which was manually evaluated by domain experts, therapists, and educators based on how well the story aligned with the selected topic or behavioral skill. The image generation component was assessed for **visual relevance and consistency**, particularly how accurately the visuals represented the story and whether the character appeared consistent across all slides. **Audio clarity and pronunciation** were judged by listening to the generated voice outputs in multiple languages, ensuring that children could understand the story narration clearly. Additionally, **system performance** was measured in terms of response time for generating each content type. The **User Satisfaction Score**, collected through feedback forms, helped assess the overall usability and emotional impact of the system. Another important parameter was the **Quiz Engagement Rate**, which captured how many users attempted and completed the quiz at the end of the story, offering insights into content retention.

7.3 Input Parameters / Features Considered

The input parameters were designed to give users flexibility in generating personalized stories that cater to individual children's needs. The most critical input was the **story topic or title**, which guided the language model to generate the narrative. In addition, users could specify the **target behavior**—such as brushing teeth, making friends, or waiting patiently—which influenced the tone and direction of the story. **Language preference** was another major input parameter, allowing users to generate outputs in their native language to make the content more accessible. The system currently supports multiple Indian and global languages for both text and audio. Although optional in the current prototype, a feature to select the **character type** (e.g., boy, girl, or cartoon animal) is being planned to enhance personalization in visual outputs. Lastly, users could specify the **output format**, choosing between text-only stories, text with images, or the full multimedia experience including audio, video, and quiz—making the system flexible for various use cases, from classrooms to therapy sessions.

7.4 Graphical and Statistical Output

To better interpret the system's performance, the outputs were visualized using statistical and graphical representations. A majority of stories were rated between **4 and 5 on a scale of 5** for text relevance, indicating a high degree of alignment with the input topic and expected social learning outcomes. The **average generation time** for each component was also recorded—text took approximately 4 seconds, each image around 6.5 seconds, audio about 3.8 seconds, and the complete video around 8 seconds. This confirmed the system's efficiency in generating full-fledged multimedia stories in under a minute. The **quiz completion rate** was recorded at 82%, with users scoring an average of 4.1 out of 5, suggesting that children retained most of the story content. These metrics validate the effectiveness of the system as a learning aid and its practicality for real-world use. The data also helped identify areas for optimization, such as reducing latency in image generation and improving character consistency across visuals.

7.5 Comparison of Results with Existing Systems

Compared to manual story creation or using multiple standalone tools like Canva for visuals and Google TTS for audio, the proposed system provides a far more efficient and unified experience. Traditional methods often require **1–2 hours** to create a complete story, involving script writing, designing visuals, voiceover recording, and formatting—all of which demand technical skill. Even semi-automated tools still require multiple steps and software tools, which can be overwhelming for parents or educators without a design background. In contrast, the proposed system completes the entire process in **less than 2 minutes** with just a few input fields, making it ideal for quick content creation in classroom or therapy settings. Additionally, **multi-language support** and built-in **quizzes** set this system apart from existing platforms. Although manual tools offer consistent visuals, they lack the AI-driven personalization and automation this system provides. Thus, the system outperforms in terms of accessibility, speed, usability, and educational value, especially in resource-constrained environments.

7.6 Inference Drawn

Based on the collected results and feedback, it can be confidently inferred that the system fulfills its core objective of creating personalized, accessible, and engaging social stories for children with ASD. The integration of advanced technologies—such as large language models for text, AI-based image generation, TTS for multilingual audio, and Streamlit for a

responsive frontend—enables a seamless and efficient workflow for both content creators and users. The high relevance scores, efficient processing times, and strong user engagement validate the system's real-world utility. Additionally, the feedback from educators and parents highlighted how this tool reduces the workload involved in preparing individualized learning materials. While challenges like API dependency and visual consistency remain, they are outweighed by the system's strengths in customization, automation, and educational impact. Overall, the system demonstrates how AI can be effectively harnessed to make inclusive education more practical, scalable, and child-centered.

Chapter 8: Conclusion

8.1 Limitations

While the proposed system demonstrates significant potential in enhancing the way social stories are created for children with Autism Spectrum Disorder (ASD), it is not without certain limitations. One primary limitation is the **dependency on third-party APIs** for core functionalities such as text generation (LLMs), image creation, and text-to-speech conversion. These APIs often come with usage limitations, quota restrictions, or associated costs, which can affect the scalability of the system in educational institutions or therapy centers with limited budgets.

Another limitation is the **inconsistency in visual outputs** from generative image models. Since image generation tools like DALL·E or Stable Diffusion may produce variations in character appearance, maintaining a consistent character throughout a multi-image story remains a technical challenge. This inconsistency can impact the clarity and coherence of the story for the child.

Moreover, the **system relies on internet connectivity** to communicate with cloud-based services. This restricts its usability in remote or underdeveloped areas with limited or unreliable internet access. Also, current multilingual support for Indian languages, especially for high-quality text-to-speech audio, is limited, which reduces accessibility for non-English-speaking users.

Finally, as the system is built entirely on **Streamlit**, its current structure leans more towards rapid prototyping than production-level robustness. For real-time classrooms or high-traffic use, further optimization and migration to a more scalable framework may be required.

8.2 Conclusion

The project successfully presents a **novel AI-driven approach** for generating personalized, multimedia-rich social stories that cater specifically to children with Autism Spectrum Disorder. By integrating technologies such as **Streamlit** (for frontend interface), **Python** (for backend logic), **Firebase** (for authentication), and **Large Language Model APIs** (for story text generation), the system automates and streamlines what has traditionally been a manual and time-consuming process.

This solution not only creates custom text stories based on user input but also enhances them with **images, audio narration, video output, and interactive quizzes**, providing a holistic, multi-sensory learning experience. It addresses a key gap in special education by allowing parents, therapists, and educators to generate tailored content that resonates with each child's unique behavioral and social learning needs.

Through this project, we demonstrate the **transformative potential of Generative AI in inclusive education**, showcasing how automation and personalization can bring meaningful improvements to how social skills are taught to neurodiverse children. The system is user-friendly, accessible, and adaptable, making it a valuable tool in the broader mission to make education more inclusive and empathetic.

8.3 Future Scope

There are several promising directions for enhancing and expanding the system in the future:

1. **Character Consistency in Image Generation**: By fine-tuning custom image generation models or using advanced prompt engineering, future versions can maintain consistent character visuals throughout the story, improving narrative continuity and child comprehension.

2. **Mobile App Integration**: Migrating or complementing the existing Streamlit web app with a dedicated mobile app (built using Flutter or React Native) can improve accessibility and offline usability, especially in low-infrastructure environments.

3. **Offline Mode**: Future development could incorporate caching mechanisms and on-device generation (using distilled LLMs or local TTS models) to support offline

use in schools without internet access.

4. **Expanded Multilingual Support**: Enhancing text and audio generation to support more Indian and global languages would broaden the reach of the tool. Incorporating high-quality regional voice synthesis would make the system more inclusive.

5. **Emotionally Adaptive Content**: Incorporating sentiment analysis and emotional feedback could enable the system to adjust stories based on the emotional state of the child, resulting in more empathetic and effective storytelling.

6. **Teacher/Admin Dashboard**: A backend panel for educators and therapists to manage student profiles, track quiz scores, and monitor learning progress would help integrate this system more fully into structured special education environments.

7. **Gamification & Interactivity**: Adding game-based interactions, drag-and-drop activities, and branching storylines can further engage children and make the learning experience even more immersive.

In summary, while the current version of the system achieves its primary objectives, there remains vast potential to scale its impact, enhance its intelligence, and integrate it more deeply into educational ecosystems for children with ASD and other learning needs.

References

Conference Proceedings:

[1] Alabdulrahman, M., Khayyat, R., Almowallad, K., and Alharz, Z., 2024, “Sarid: Arabic Storyteller Using a Fine-Tuned LLM and Text-to-Image Generation,” *Proceedings of the 16th International Conference on Computer and Automation Engineering (ICCAE)*.

Journal Papers / Technical Reports:

[2] Gou, C., Felemban, A., Khan, F. F., Zhu, D., Cai, J., Rezatofighi, H., and Elhoseiny, M., 2024, “How Well Can Vision Language Models See Image Details?,” *arXiv preprint*, arXiv:2408.03940.

[3] Shen, J., Mire, J., Park, H. W., Breazeal, C., and Sap, M., 2024, “HEART-felt Narratives: Tracing Empathy and Narrative Style in Personal Stories with LLMs,” *arXiv preprint*, arXiv:2405.17633.

[4] Chang, H., Zhang, H., Barber, J., Maschinot, A. J., Lezama, J., Lu, J., Yang, M.-H., Murphy, K., Freeman, W. T., Rubinstein, M., Li, Y., and Krishnan, D., 2023, “Muse: Text-To-Image Generation via Masked Generative Transformers,” *arXiv preprint*, arXiv:2301.00704.

[5] Anonymous, 2024, “Cross-Lingual Conversational Speech Summarization with Large Language Models,” *arXiv preprint*, arXiv:2408.06484.

[6] Anonymous, 2024, “CT-Eval: Benchmarking Chinese Text-to-Table Performance in Large Language Models,” *arXiv preprint*, arXiv:2405.12174.

[7] Anonymous, 2024, “LLM-DetectAIve: A Tool for Fine-Grained Machine-Generated Text Detection,” *arXiv preprint*, arXiv:2408.04284v1.

[8] Anonymous, 2024, “LLaVA-Surg: Towards Multimodal Surgical Assistant via Structured Surgical Video Learning,” *arXiv preprint*, arXiv:2408.07981.

[9] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, L., and Chen, W., 2021, “LoRA: Low-Rank Adaptation of Large Language Models,” *arXiv preprint*, arXiv:2106.09685.

[10] Khanuja, S., Dandapat, S., and Bhattacharyya, P., 2021, “MuRIL: Multilingual Representations for Indian Languages,” *arXiv preprint*, arXiv:2103.10730.