

Generating AI-driven Interactive Bedtime Stories Based on a Child's Mood and Preferences

Vani K S, Rakshith D N, Rohit Harischandra Naik, Rakshith B G, Sai Kumar C, Ojes P Dore

Nitte (Deemed to be University), Nitte Meenakshi Institute of Technology (NMIT),

Department of Information Science and Engineering,

Bangalore 560064, Karnataka, India,

Abstract

This paper presents an AI-powered interactive storytelling system that generates personalized bedtime stories tailored to a child's emotional state and individual preferences. Traditional storytelling methods, while comforting, are static and fail to adapt to a child's fluctuating moods, potentially limiting emotional engagement and therapeutic potential. Our system addresses this limitation by integrating emotion detection, narrative generation, and emotionally expressive speech synthesis into a unified framework. We employ a fine-tuned BERT-based model to identify the child's emotional tone from text inputs, which then informs prompt construction for large language models—specifically, LLaMA fine-tuned with Low-Rank Adaptation (LoRA) or Gemini 1.5 Pro—to generate mood-aligned storylines. These narratives are subsequently converted into natural, emotionally resonant speech using cutting-edge TTS systems like XTTS or Eleven Labs. A web-based interface, built with React and Flask, allows caregivers to input mood cues and instantly access AI-generated storytelling experiences in real time. Experimental results show a classification accuracy of 75.4

Keywords: Emotion detection, storytelling, LoRA, LLaMA, bedtime stories, NLP, BERT, TTS.

1. Introduction

Bedtime stories have long stood as a cherished ritual in homes across the globe—offering not just entertainment, but a sense of comfort, imaginative escape, and emotional closeness between children and their caregivers. Yet despite their cultural significance, traditional storytelling remains largely static: pre-written narratives that are indifferent to a child's shifting emotional world. In a time where digital systems increasingly adapt to user preferences and contexts, storytelling—especially for young listeners—has remained notably untouched by responsive, affective technologies.

This paper aims to address this disconnect by designing an AI-powered, emotionally adaptive storytelling platform that evolves with a child's emotional state. Rooted in the belief that stories can do more than entertain—that they can soothe, uplift, and heal—we propose a system that generates personalized, emotionally attuned bedtime stories in real time.

The architecture integrates four key components: (1) emotion detection, using a BERT-based classifier trained on annotated emotional dialogue data;

(2) story generation, powered by fine-tuned LLaMA or Gemini models that craft narratives aligned with the detected mood;

(3) expressive speech synthesis, using advanced text-to-speech systems like XTTS or Eleven Labs to deliver human-like emotional intonation; and

(4) a user-friendly web interface, enabling caregivers to either manually input or automatically detect a child's mood to initiate the story session.

By fusing natural language understanding, generative modelling, and affective computing, our system moves beyond rigid storytelling formats. A child experiencing anxiety, for example, can receive a calming and reassuring tale tailored to their emotional state—transforming story time into an empathetic, adaptive experience. This work contributes to the broader field of affective AI and child-centric technology design, highlighting how artificial intelligence can deepen—not dilute—human connection. In doing so, it reimagines

bedtime storytelling as a dynamic dialogue between emotion, imagination, and technology.

The remainder of this paper is structured as follows. Section 2 reviews related work on emotion-aware storytelling and recent advances in AI-driven personalization for children’s emotional well-being. Section 3 details the proposed system architecture, including modules for emotion detection, dynamic story generation, expressive text-to-speech synthesis, and user interface design. Section 4 presents the implementation setup and key experimental results, evaluating model performance across multiple metrics. Section 5 and 6 talks about the limitations of the system and the possible future work in this area. Finally, Section 7 concludes with a summary of the whole research work.

2. Objectives

1. To design an AI-powered interactive storytelling framework that dynamically generates personalized bedtime stories tailored to a child’s emotional state and individual preferences.
2. To develop an emotion detection module using a fine-tuned BERT-based model that accurately identifies a child’s emotional tone from textual inputs to guide story personalization.
3. To generate mood-aligned and emotionally coherent narratives by integrating emotion-aware prompt engineering with large language models such as LLaMA (with LoRA fine-tuning) or Gemini 1.5 Pro.
4. To enhance emotional engagement through expressive speech synthesis by converting generated narratives into natural, emotionally resonant audio using advanced TTS systems like XTTS and Eleven Labs, delivered via a real-time web-based interface

The primary objective of the proposed system is to create an AI-driven interactive storytelling platform that delivers personalized bedtime stories based on a child’s emotional state and preferences. By incorporating a fine-tuned BERT-based emotion detection model, the system accurately identifies the emotional tone from user-provided text inputs. This emotional context is used to guide prompt construction for large language models, enabling the generation of mood-aligned and emotionally coherent narratives. The system leverages

advanced LLMs such as LLaMA with LoRA fine-tuning or Gemini 1.5 Pro to ensure high-quality and contextually relevant story generation. To further enhance emotional engagement, the generated stories are transformed into expressive and natural speech using state-of-the-art TTS systems like XTTS or Eleven Labs. Finally, a web-based interface built using React and Flask enables real-time interaction, allowing caregivers to easily access and personalize storytelling experiences.

3. Methods

The proposed system is an AI-driven storytelling pipeline that dynamically generates and narrates personalized bedtime stories based on a child’s mood and preferences. The solution comprises four main modules: emotion detection, story generation, emotional text-to-speech, and a web-based user interface.

The entire methodology is given in Fig 1

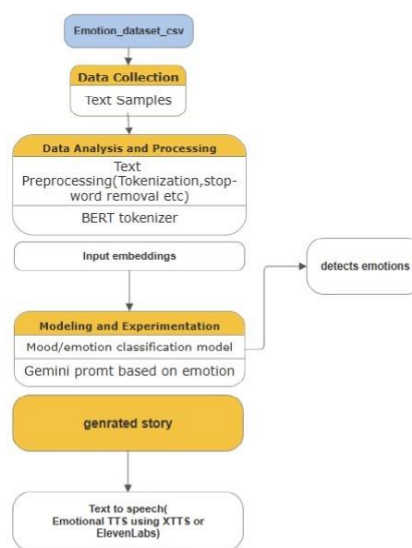


Fig. 1: System pipeline from emotion detection to story playback

The entire Use Case Diagram for the Bed Time Story Generation is given in Fig 2

A. Data Collection

- 1) Emotion Detection Dataset: We utilized the emotion dataset raw.csv file from Kaggle, which contains 34,795 text entries labeled with eight emotions: anger, disgust, fear, joy, neutral, sadness,

shame, and surprise. The data, primarily from tweets, offers diverse emotional expressions.

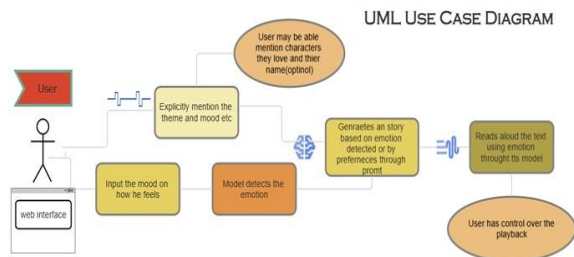


Fig. 2: Use Case Diagram for Bed Time Story Generation

2) Story Generation Dataset: A dataset from Hugging Face, consisting of 100 paired entries of story prompts and corresponding short stories, was used to fine-tune LLaMA using LoRA (Low-Rank Adaptation), ensuring better narrative alignment with emotional contexts.

B. Data Preprocessing

- Tokenization and Lowercasing: Implemented using NLTK
- Stop-word Removal and Lemmatization: Using Word-NetLemmatizer
- Text Cleaning: Removal of non-alphabetic characters
- Label Encoding: Emotion labels numerically encoded using Scikit-learn's LabelEncoder
- BERT Embedding: Text embedded using bert-base-uncased tokenizer
- Train-Test Split: Conducted using train test split from Scikit-learn

C. Model Architecture and Components

1) Emotion Detection: • Model: Fine-tuned bert-base-uncased

- Frameworks: PyTorch and Hugging Face Transformers
- Training Setup:
 - Optimizer: AdamW
 - Loss: Cross Entropy Loss
 - Epochs: 5
 - Batch Size: 16

– Learning Rate: 5×10^{-5}

- Evaluation: Accuracy, classification report, and confusion matrix

2) Story Generation:

- Gemini 1.5 Pro: Google's state-of-the-art multimodal model for cloud-based generation, offering superior narrative coherence and emotional alignment via API.

- Primary Model: LLaMA (Meta) fine-tuned using LoRA

- Fine-tuning Method: Low-Rank Adaptation (LoRA) for parameter efficiency

- Prompting: Dynamic prompt construction using detected emotion and preferences

3) Text-to-Speech:

- Tools: XTTS v2 (Coqui) or Eleven Labs Multilingual v2

- API Integration: Eleven Labs requires xi-api-key

4) User Interface: • Frameworks: React + Vite + TypeScript + Flask

- Features: Mood input, story display, audio playback, feedback capture

- Designed for ease of use by caregivers, with a minimal and responsive layout

- Real-time interaction between frontend and backend ensures smooth story generation

- Includes error handling and visual loading indicators to improve usability

D. System Flow

The story generation pipeline follows:

User Input → Emotion Detection → Prompt Generation → Story Generation (LLaMA) → Emotional TTS → Web Playback

The system begins by receiving user input, which is first analysed to detect the user's emotional state. Based on the detected emotion, a context-aware prompt is automatically generated to guide the story flow. This prompt is then fed into a LLaMA-based language model to generate a personalized story. The generated story is converted into emotionally expressive speech using an emotional

Text-to-Speech (TTS) module. Finally, the synthesized audio is delivered to the user through a web-based playback interface.

3. Results

A. Interpretation of Results

A. Emotion Detection Performance

Fig. 3 shows that the emotion classifier achieved 75.40% overall accuracy with varying performance across emotions:

TABLE I
EMOTION CLASSIFICATION PERFORMANCE

Emotion	Precision	Recall	F1-score
Anger	0.65	0.41	0.50
Disgust	0.00	0.00	0.00
Fear	0.64	0.64	0.64
Joy	0.69	0.76	0.72
Neutral	0.78	0.70	0.74
Sadness	0.33	0.67	0.44
Shame	0.00	0.00	0.00
Surprise	0.25	0.08	0.12
Accuracy			0.75

Fig. 3: Classification Report

Fig. 4 shows the confusion matrix of the trained model.

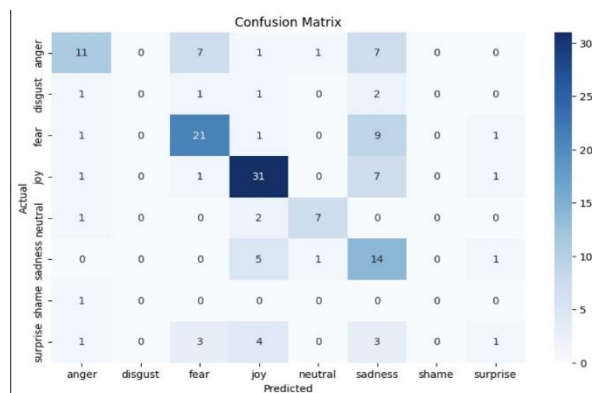


Fig. 4: Confusion Matrix

B. Story Generation Samples

- Input Emotion: Joy

Generated Story (LLaMA): "The sun was shining brightly as Alice skipped down the path. She had just gotten her dream job and couldn't wait to celebrate with her friends." Generated Story (Gemini 1.5 Pro): "Laughter filled the air as Mia danced in the rain, her yellow boots splashing puddles. Today was perfect—she had adopted a puppy named Sunny!"

- Input Emotion: Sadness

Generated Story (LLaMA): "The room was silent. His favorite book lay closed on the desk. The memories rushed back, each more painful than the last." Generated Story (Gemini 1.5 Pro): "The old treehouse stood empty, its ropes frayed. Emma traced her fingers over the carved initials, remembering how her brother used to read stories there."

Fig. 5 shows the UI which is used to generate the bedtime story along with additional playback view

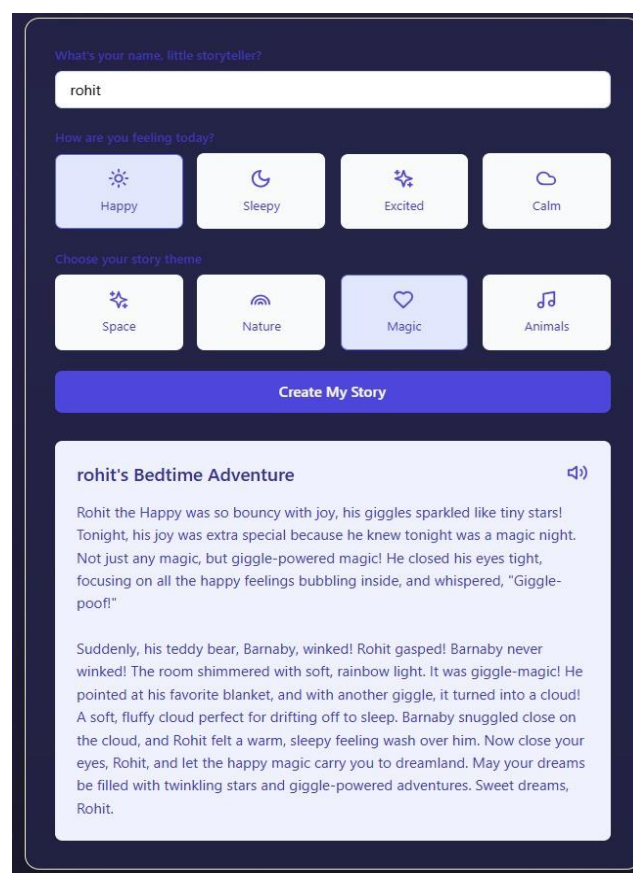


Fig. 5: User Interface used to generate story

4. Discussion

Key Observations

- Strong performance for high-frequency emotions (joy, neutral, fear) with F1-scores ≥ 0.64
- Poor performance on minority classes (disgust, shame, surprise) due to class imbalance
- LLaMA fine-tuned with LoRA produced more emotionally consistent stories than base models

Emotionally-conditioned prompts for common emotions yielded coherent stories

- Stories for sadness/surprise sometimes showed emotional ambiguity

Limitation

Class Imbalance: Poor performance on minority emotion classes (disgust, shame)

- Computational Costs: Fine-tuning LLaMA with LoRA requires significant resources
- PromptSensitivity: Output quality varies with prompt wording
- Emotion Depth: Generated stories sometimes lack nuanced emotional progression.
- Gemini 1.5 Pro requires API access while LLaMA can run offline

Future Work

- Address class imbalance using synthetic data generation (e.g., GPT-based augmentation)
- Implement reinforcement learning from human feedback (RLHF) for LLaMA
- Develop multi-modal emotion detection (voice tone, facial expressions)
- Optimize LoRA fine-tuning for better emotional consistency
- Add interactive story branching based on real-time child feedback

Fig. 6 shows the UI which is used to generate the bedtime story

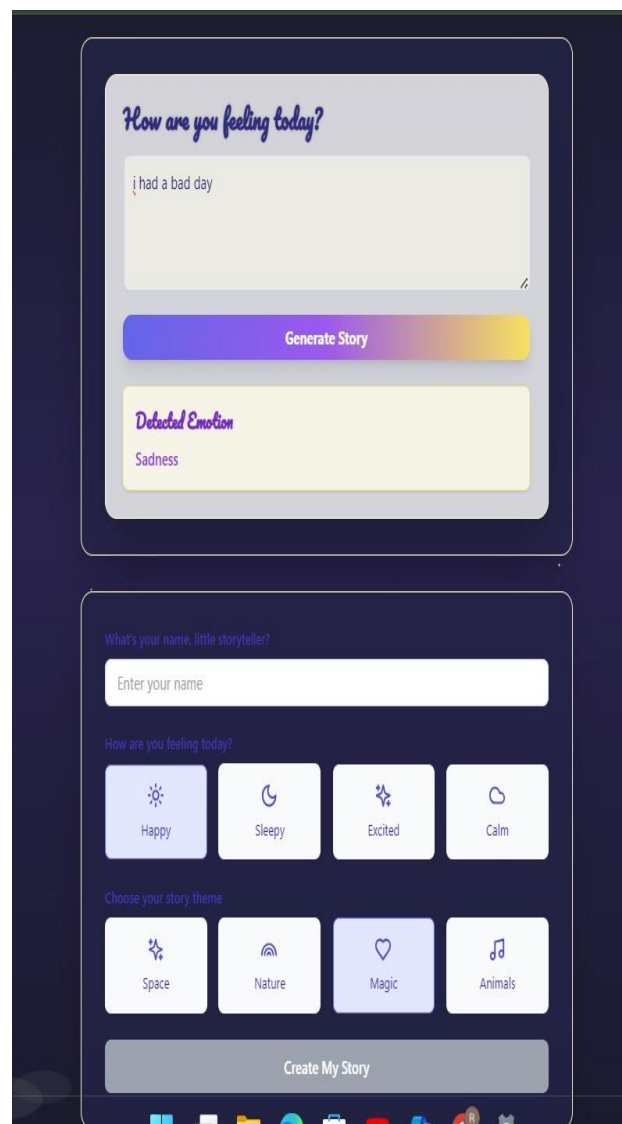


Fig. 6: User Interface used to generate story

5. Conclusion

This paper presented an AI-driven system that transforms traditional bedtime storytelling into personalized, emotionally intelligent experiences for children. By integrating emotion detection (fine-tuned BERT), dynamic story generation (LLaMA fine-tuned with LoRA), and expressive text-to-speech (XTTS/Eleven Labs), the system creates context-aware narratives. The React-based web interface enables caregivers to easily interact with the system.

Key achievements include 75.4 emotion classification accuracy and emotionally consistent story generation through LoRA fine-tuning. Future work will focus on multi-modal emotion detection

and interactive story branching. This project highlights the potential of parameter-efficient fine-tuning (PEFT) methods like LoRA in creating emotionally aware AI systems for children's well-being.

References

- [1] Chubb, J., Missaoui, S., Concannon, S., Maloney, L., & Walker, J. A. (2021). Interactive storytelling for children: A case-study of design and development considerations for ethical conversational AI. *International Journal of Child-Computer Interaction*, 32, 100403. <https://doi.org/10.1016/j.ijcci.2021.100403>
- [2] K.S.Kolekar, D.J.Richter, M.I.Bappi, and K.Kim, "Advancing AI Voice Synthesis: Integrating Emotional Expression in Multi-Speaker Voice Generation," in Proc. IEEE Int. Conf. AI Intell. Inform. Comput. (ICAIIIC), Gwangju, South Korea, pp.193–198, 2024.
- [3] Sale, H. B., Ghodake, A. P., Indalkar, M. M., Gamare, M. M., Jagtap, S. S., & Shirsat, S. M. (2024). AI Bedtime Storyteller using Chatgpt Prompt. 2022 6th International Conference on Computing, Communication, Control and Automation (ICCUBEA), 1–5. <https://doi.org/10.1109/iccubea61740.2024.10775073>
- [4] B.Burtenshaw, "AI Stories: An Interactive Narrative System for Children," arXiv preprint, Nov.9,2020
- [5] Konarivand, N. H. (2025). AI-Driven storytelling and its impact on discipline and engagement in TaekFunDo. *Journal of Information Systems Engineering & Management*, 10(17s), 383–394. <https://doi.org/10.52783/jisem.v10i17s.2744>
- [6] Steinhäusser, S. C., Zehe, A., Schnetter, P., Hotho, A., & Lugin, B. (2024). Towards the development of an automated robotic storyteller: comparing approaches for emotional story annotation for non-verbal expression via body language. *Journal on Multimodal User Interfaces*. <https://doi.org/10.1007/s12193-024-00429-w>
- [7] T. Vijayakumaran, T. Thavachelvam, A. Gnaeswaran, T. C. Sandanayake, K. A. S. N. Sumathipala, and S. Bandara, "Development of Digital Storytelling Platform for Children Based on Emotions," in 2021 6th International Conference on Information Technology Research (ICITR), Moratuwa, Sri Lanka, Dec.2021, pp.1–6. doi:10.1109/ICITR54349.2021.9657431
- [8] Y.-X. Li and N.-C. Tai, "Teaching at the Right Moment: A Generative AI-Enabled Bedtime Storybook Generation System Communicating Timely Issues," in Proceedings of the 2023 International Conference on Consumer Electronics – Taiwan (ICCE-Taiwan 2023), Pingtung, Taiwan, Jul.2023, pp.157–158. Institute of Electrical and Electronics Engineers Inc.
- [9] A. A. Schlarb and J. Faber, "Artificial intelligence (AI) in pediatric sleep: AI vs. expert-generated psychotherapeutic pediatric sleep stories — Can reviewers detect the difference?," *Somnologie*, accepted 5 December 2024, published 17 December 2024, vol. –, no. –, pp. –, doi:10.1007/s11818-024-00491-2
- [10] P.Nandwani and R.Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol.11, no.1, p.81, Aug.2021, doi:10.1007/s13278-021-00776-6.