IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

AI-Based KidsLLM System for Secure Educational Assistance : A Research Study on Content Filtering, Ethical Response Mechanism, and Child-Safe Conversational Interfaces

Trupti Suresh Sakpal[1] and Sandeep Kumar Vishwakarma[2]
Student, MCA[1]
Head, Department of Information Technology[2]
Chandrabhan Sharma College of Arts Commerce and Science, Mumbai, India
truptisakpal24@gmail.com and sandeepvcbs@gmail.com

Abstract: *Abstract This research paper presents the design, implementation, and evaluation of an AI-Based KidsLLM system developed to provide secure educational assistance for children through safe conversational artificial intelligence. The proposed system integrates local large language models, ethical response mechanisms, content filtering algorithms, and child-safe conversational interfaces to ensure educationally appropriate interactions.*

The study focuses on the implementation of Flask-based web architecture integrated with Ollama LLM, MongoDB database systems, user authentication modules, and real-time harmful content moderation. The research explores system architecture, data collection, feature extraction, model training, evaluation, optimization, and recommendation generation.

Experimental analysis demonstrates the effectiveness of KidsLLM in filtering harmful prompts, maintaining child-safe communication, and improving educational interaction quality. The proposed framework contributes toward ethical AI implementation for children and supports future advancements in responsible conversational systems.

KidsLLM is a child-friendly conversational artificial intelligence system designed to provide safe, filtered, and educational interactions for children. The system uses local Large Language Models (LLMs) integrated with Python, Flask, MongoDB, and content moderation layers to ensure secure communication.

Unlike traditional AI chat systems that may expose children to harmful language or unsafe topics, KidsLLM focuses on filtered responses, educational assistance, and age-appropriate interactions. This research journal discusses the architecture, methodology, database design, implementation, testing, security measures, and future scope of the KidsLLM system.

Keywords: *KidsLLM.*

I. INTRODUCTION

Artificial Intelligence has transformed modern communication systems through intelligent chatbots and virtual assistants. However, most public AI systems are not specifically designed for children. KidsLLM addresses this issue by introducing a safe and educational AI platform for kids. The project combines local LLM technology with safety filtering, keyword moderation, and secure backend development. The system uses Flask for the web framework, MongoDB for database storage, and Ollama with Llama models for local AI execution. The objective is to create a controlled environment where children can learn, ask questions, and complete school-related activities

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, May 2026 Impact Factor: 8.2

without exposure to harmful content. The research also focuses on responsible AI implementation, ethical content filtering, and secure communication between the frontend and backend systems. Artificial Intelligence has transformed modern communication systems through intelligent chatbots and virtual assistants. However, most public AI systems are not specifically designed for children.

KidsLLM addresses this issue by introducing a safe and educational AI platform for kids. The project combines local LLM technology with safety filtering, keyword moderation, and secure backend development. The system uses Flask for the web framework, MongoDB for database storage, and Ollama with Llama models for local AI execution. The objective is to create a controlled environment where children can learn, ask questions, and complete school-related activities without exposure to harmful content. The research also focuses on responsible AI implementation, ethical content filtering, and secure communication between the frontend and backend systems. Artificial Intelligence has transformed modern communication systems through intelligent chatbots and virtual assistants. However, most public AI systems are not specifically designed for children.

KidsLLM addresses this issue by introducing a safe and educational AI platform for kids. The project combines local LLM technology with safety filtering, keyword moderation, and secure backend development. The system uses Flask for the web framework, MongoDB for database storage, and Ollama with Llama models for local AI execution. The objective is to create a controlled environment where children can learn, ask questions, and complete school-related activities without exposure to harmful content. The research also focuses on responsible AI implementation, ethical content filtering, and secure communication between the frontend and backend systems.

II. LITERATURE REVIEW

Existing studies in conversational AI highlight the rapid development of intelligent systems powered by machine learning and natural language processing. Research on educational chatbots shows that AI can improve learning experiences by providing instant support and interactive guidance. However, researchers have identified major risks when children use unrestricted AI systems. Unsafe language, misinformation, and privacy concerns remain key challenges. Several educational AI systems use moderation APIs, but cloud dependency and internet privacy issues are still concerns. KidsLLM contributes to this field by introducing a locally hosted AI architecture combined with safety filtering and child-focused content moderation. The proposed model reduces privacy risks and ensures safer educational interactions. Existing studies in conversational AI highlight the rapid development of intelligent systems powered by machine learning and natural language processing. Research on educational chatbots shows that AI can improve learning experiences by providing instant support and interactive guidance.

However, researchers have identified major risks when children use unrestricted AI systems. Unsafe language, misinformation, and privacy concerns remain key challenges. Several educational AI systems use moderation APIs, but cloud dependency and internet privacy issues are still concerns. KidsLLM contributes to this field by introducing a locally hosted AI architecture combined with safety filtering and child-focused content moderation. The proposed model reduces privacy risks and ensures safer educational interactions. Existing studies in conversational AI highlight the rapid development of intelligent systems powered by machine learning and natural language processing. Research on educational chatbots shows that AI can improve learning experiences by providing instant support and interactive guidance. However, researchers have identified major risks when children use unrestricted AI systems. Unsafe language, misinformation, and privacy concerns remain key challenges. Several educational AI systems use moderation APIs, but cloud dependency and internet privacy issues are still concerns. KidsLLM contributes to this field by introducing a locally hosted AI architecture combined with safety filtering and child-focused content moderation. The proposed model reduces privacy risks and ensures safer educational interactions.Objectives of KidsLLM Project

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, May 2026 Impact Factor: 8.2

Primary Objectives :

1. To develop a child-safe conversational AI system that provides secure, educational, and age-appropriate responses for children using Large Language Models (LLMs).
2. To implement intelligent content filtering mechanisms capable of detecting and blocking harmful, toxic, violent, abusive, adult, or misleading content before it reaches child users.
3. To design an ethical response generation framework that ensures AI-generated responses follow child-friendly communication standards and educational values.
4. To create a secure educational assistance platform using technologies such as Python, Flask, Ollama, and MongoDB for interactive learning support.
5. To provide a safe digital learning environment where children can ask academic, scientific, and general knowledge questions without exposure to unsafe internet content.
6. To integrate real-time moderation and prompt inspection systems for monitoring user interactions and maintaining conversational safety.
7. To develop administrative control modules that allow administrators to inspect prompts, authorize registrations, block suspicious activities, and monitor system behavior.
8. To maintain user privacy and secure data management through encrypted authentication systems and protected database storage.
9. To evaluate the effectiveness of AI moderation techniques in reducing unsafe responses and improving educational interaction quality.
10. To promote responsible and ethical AI usage in education by implementing safe conversational interfaces specifically designed for children.

III. EXISTING SYSTEM

Traditional chatbot systems such as general AI assistants are primarily designed for adults and unrestricted conversations. Most systems rely heavily on cloud infrastructure and internet-based APIs. While these systems provide accurate responses, they often lack strict child safety controls.

The limitations of existing systems include :

• Exposure to harmful or inappropriate language.
• Lack of educational filtering.
• Internet dependency and privacy risks.
• High operational costs for API usage.
• Limited parental control features.

These challenges justify the need for a child-friendly AI solution such as KidsLLM. Traditional chatbot systems such as general AI assistants are primarily designed for adults and unrestricted conversations. Most systems rely heavily on cloud infrastructure and internet-based APIs. While these systems provide accurate responses, they often lack strict child safety controls. The limitations of existing systems include: • Exposure to harmful or inappropriate language.

Scope of the Study:

The scope of the KidsLLM project focuses on the development and implementation of a secure AI-powered educational chatbot system designed specifically for children. The project aims to combine conversational artificial intelligence, ethical AI principles, and content moderation techniques to create a safe educational platform.

The system includes:

- Child-safe conversational interfaces
- Educational query handling

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

- Real-time harmful content filtering
- AI-generated educational responses
- User authentication and registration systems
- MongoDB database management
- Administrative monitoring and moderation
- Chat history management
- PDF generation of educational responses
- Secure backend architecture using Flask API
- Local LLM deployment using Ollama

IV. PROPOSED SYSTEM REVIEW :THE PROJECT MAINLY COVERS

The proposed KidsLLM system introduces a secure and educational AI chatbot designed specifically for children. The architecture uses a local Large Language Model executed through Ollama and integrated with Flask-based APIs. The proposed system contains:

• User-friendly web interface.
• Input filtering system.
• Harmful keyword detection.
• Educational response generation.
• MongoDB-based user management.
• Session handling and authentication.

1. Educational Assistance :

KidsLLM provides academic support for subjects such as science, mathematics, general knowledge, and project-based learning through conversational AI interaction.

2. AI Safety and Moderation :

The system focuses on filtering unsafe prompts and preventing harmful responses through keyword detection, moderation rules, and ethical response mechanisms.

3. Secure System Architecture :

The platform ensures secure handling of user credentials, chat histories, and moderation logs using MongoDB and authentication systems.

4. Research and Experimental Analysis :

The project evaluates the effectiveness of content filtering, response safety, user interaction quality, and system performance through empirical testing.

Overall System Architecture :

The architecture follows a client-server model where users interact through a web interface connected to Flask APIs. MongoDB stores registration details, login credentials, chat history, and administrative records. The Ollama-hosted Llama model processes educational queries and generates intelligent responses. A content filtering engine validates prompts and generated responses before displaying them to users.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, May 2026 Impact Factor: 8.2

Data Collection :
Educational prompts and child-safe datasets are collected from educational repositories and curated learning resources. Unsafe keywords including adult content, violence, drugs, and abusive language are stored in moderation dictionaries. Administrative logs and interaction records are maintained for performance evaluation.

Data Processing :
Input prompts undergo preprocessing including tokenization, stop-word removal, normalization, and keyword matching. Sensitive terms are identified using regular expression matching and semantic analysis. Processed prompts are forwarded to the LLM only after safety verification

Feature Extraction :
Feature extraction identifies educational intent, toxicity score, sentence polarity, and contextual relevance. Natural Language Processing techniques are used to classify educational and harmful prompts. Features support ethical response generation and conversational filtering.

Model Training :
The system utilizes pretrained local language models integrated through Ollama. Prompt-response datasets are evaluated to improve educational alignment. Safety prompts are embedded to guide the model toward child-safe behavior.

Recommendation System :
The recommendation module suggests educational topics and safe learning resources. User interaction history helps generate personalized educational assistance.

Evaluation and Optimization :
System evaluation includes response accuracy, safety compliance, toxicity reduction, and user satisfaction. Optimization techniques improve latency, filtering efficiency, and database performance.

V. METHODOLOGY

The development methodology follows a modular software engineering approach.The project consists of frontend, backend, database, AI integration, and safety modules.
The workflow includes:
1. The user enters a query.
2. Flask backend receives the request.
3. Safety filters analyze the content.
4. Safe requests are forwarded to the LLM.
5. AI response is generated.
6. Output is filtered again before displaying.

Python programming language was used for backend development because of its strong AI ecosystem and Flask integration support.
The research methodology follows a structured software engineering and AI development lifecycle. Educational datasets and harmful keyword datasets are collected and processed. Training and validation datasets are divided into structured subsets for evaluation.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, May 2026 Impact Factor: 8.2

1. System Architecture :The system consists of three main components:
   - Frontend: A simple web interface for user interaction
   - Backend: A Python-based server to process requests
   - Database:Mongodb
   - AI Integration : LLM Engine, A locally running language model LLma3 (e.g., via Ollama)
   - Safety modules: Flaskpymongo,ollama,werkzeug,logging,re (Regular Expressions),os,functool
   - Filtering Module: Custom Python scripts

Database Design :

MongoDB was selected because of its flexibility and compatibility with Python applications.The database stores user profiles, login credentials, and chat records. Collections used in the project:

• Users Collection
• Chat History Collection
• Moderation Logs
• Admin Settings

Each collection uses JSON-like documents that allow scalable storage and efficient retrieval. MongoDB was selected because of its flexibility and compatibility with Python applications. The database stores user profiles, login credentials, and chat records. Collections used in the project.

The project also includes a seed.py module used for database initialization and default data insertion.

Technologies Used

- Programming Language: Python
- Frameworks: Flask / FastAPI
- Frontend: HTML, CSS, JavaScript
- LLM Integration: Ollama (LLaMA 3 or similar model)

Recommended Hardware Requirements:

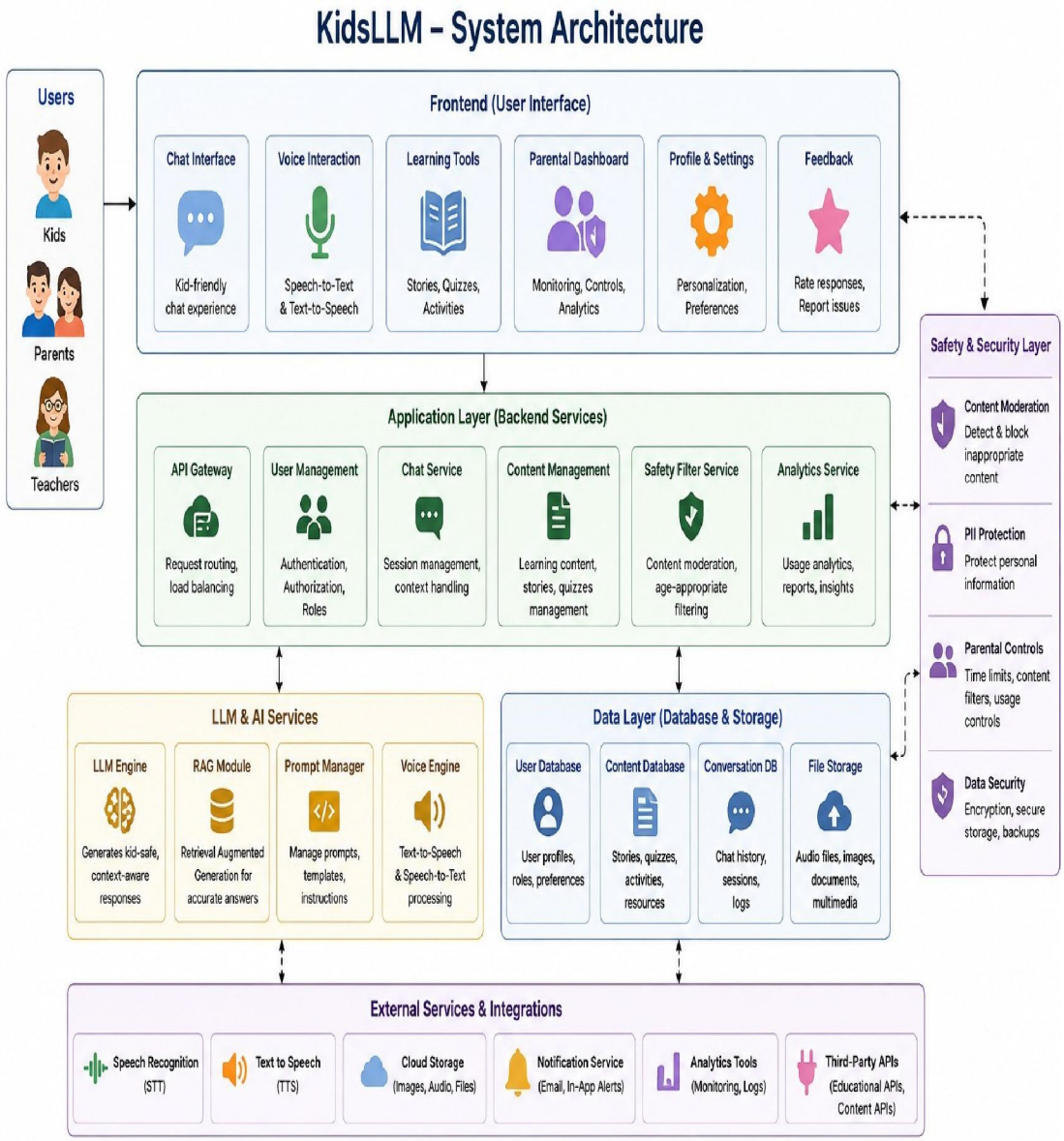
Component	Requirement
Processor	Intel i7 / Ryzen 7
RAM	16 GB or higher
Storage	512 GB SSD
GPU	NVIDIA GPU (for faster AI inference)
VRAM	6 GB or higher
Internet	High-speed broadband

Server Requirements (Optional for Deployment)

Component	Requirement
Web Server	Flask Development Server / Gunicorn
Database Server	MongoDB Server
Deployment Platform	Localhost / Cloud Server
Backup Storage	External or Cloud Backup

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

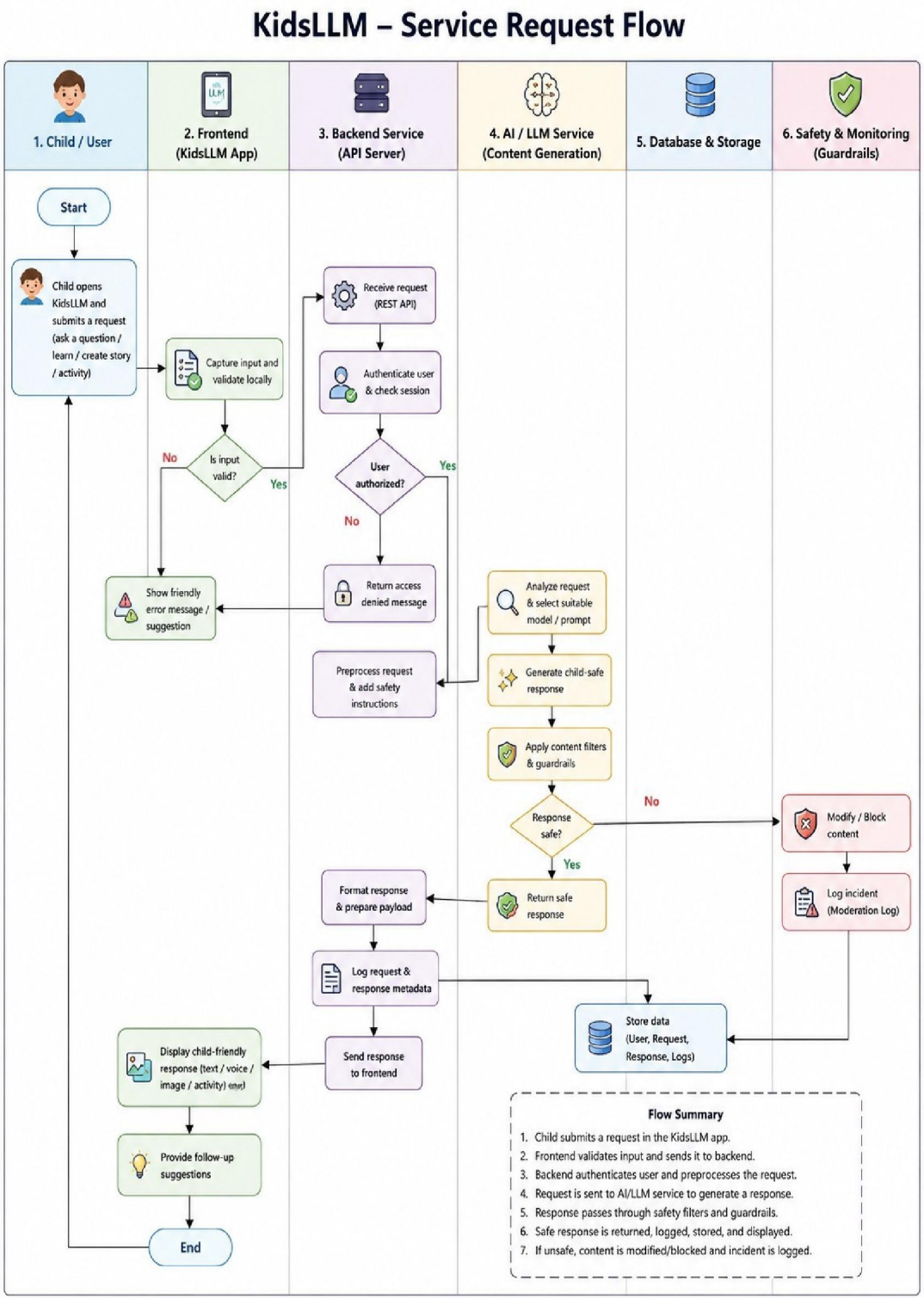
VI. SYSTEM ARCHITECTURE DIAGRAM



DOI: 10.48175/IJARSCT-35718

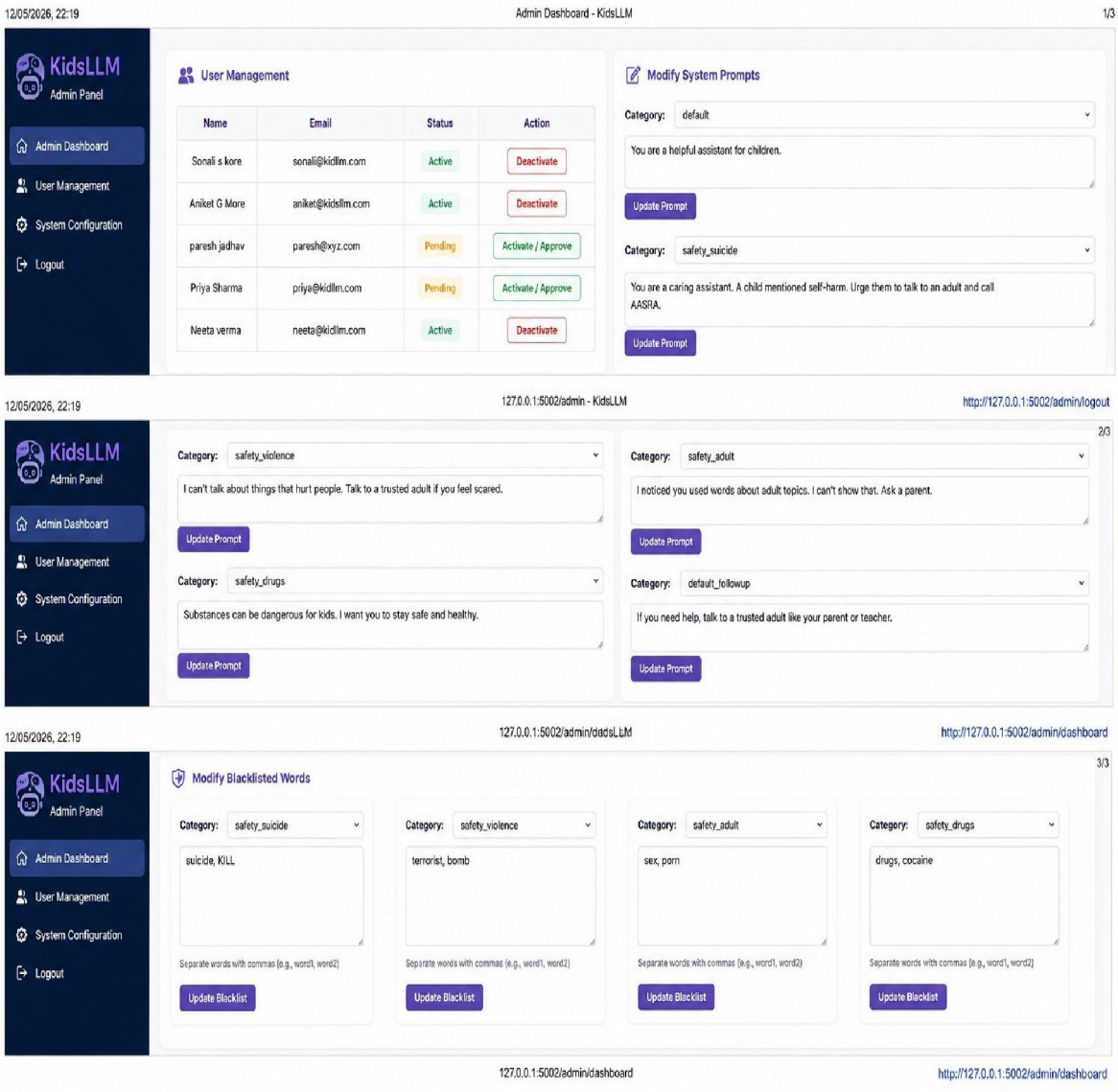
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

User Work Flow Diagram :



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

Admin Dashboard view :



Advantages :

- Protects children from harmful content
- Encourages safe learning with AI
- Works without internet dependency (local model)
- Customizable filtering rules.

Limitations within Scope:

- The system currently depends on predefined moderation rules and filtering datasets
- Advanced contextual understanding of harmful intent may require more sophisticated AI moderation

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, May 2026
Impact Factor: 8.2

models.

- Performance depends on local hardware resources when running local LLMs.
- Limited knowledge compared to cloud-based models
- Filtering may not catch all edge cases
- Requires system resources to run locally .

Future Scope:

The KidsLLM project has significant future potential in the field of educational technology and child-safe AI systems , such as multilingual support,Speech-to-text and text-to-speech technologies ,Emotion-aware AI,educational games to create a more interactive learning environment. Although the current system provides safe text-based interactions, several advanced features can be added in future versions to improve functionality, scalability, and user experience.

Conclusion:

Testing and implementation results show that KidsLLM effectively blocks harmful content while maintaining smooth conversational interaction.In conclusion, KidsLLM represents an important step toward responsible and ethical AI development for children. The project highlights how Artificial Intelligence can be safely adapted for educational use while maintaining privacy, security, and parental control.

REFERENCES

[1]. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.
[2]. Flask Official Documentation, Flask Official Website
[3]. MongoDB Official Documentation, MongoDB Official Website
[4]. Ollama Documentation, Ollama Official Website
[5]. Python Software Foundation Documentation, Python Official Website
[6]. Research papers on Natural Language Processing and Conversational AI Systems.
[7]. Studies on AI Safety and Child-Friendly Artificial Intelligence Platforms.
[8]. Educational chatbot research articles and machine learning publications.
[9]. Werkzeug Security Documentation, Werkzeug Official Website
[10]. Research journals related to Ethical AI and Responsible Computing.