

AI Generated PicStory

**“Transform Your Memories into Captivating
Visual Stories – Instantly & Effortlessly!”**



AGENDA

1.

1. Introduction
2. Problem Statement
3. Simplified Solution Diagram
4. Approach and Innovations
5. Tech Stack
6. LLM Integration & Challenges
7. Impact & Future Outlook
8. Conclusion and Team Contributions

INTRODUCTION

We capture so many moments through photos and videos, but what if we could turn them into beautiful, narrated stories effortlessly? Our AI-powered PicStory Generator does exactly that! Simply upload your images, videos, or voice notes, and let AI create a captivating visual story with automatic descriptions, translations, and voiceovers.

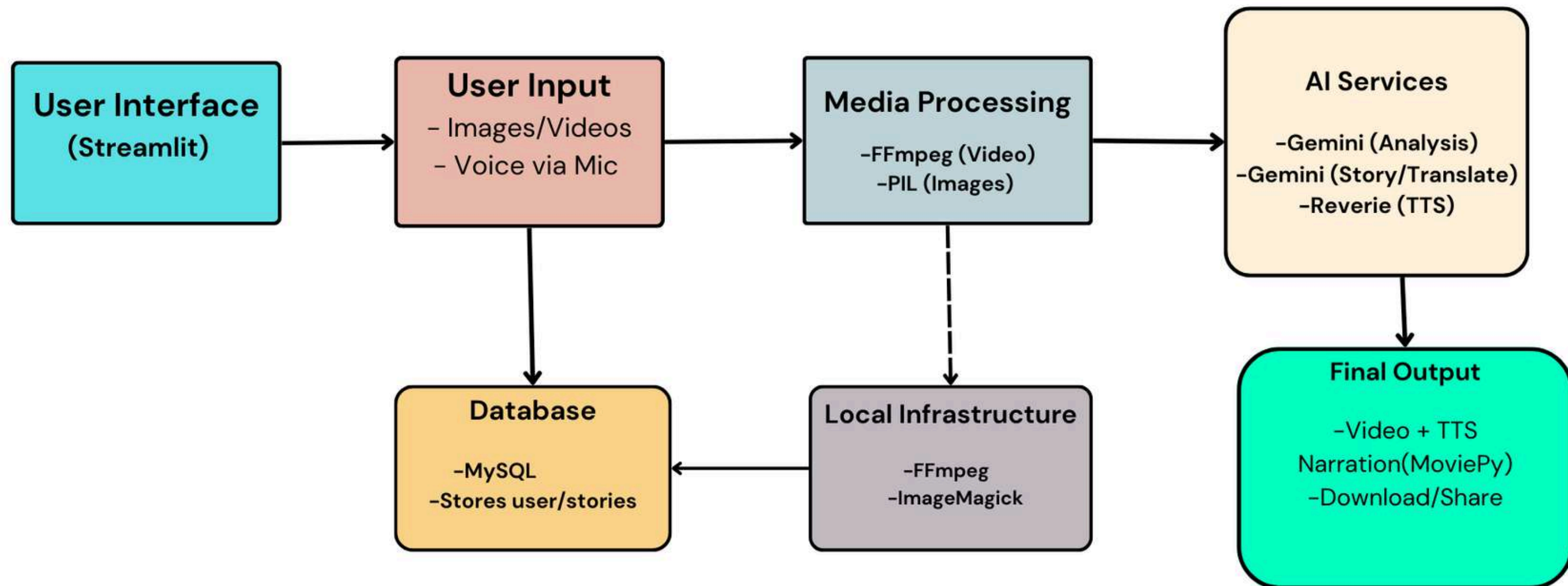
With smart media processing, real-time previews, and seamless synchronization, you can transform your memories into stunning videos in just a few clicks. Whether it's for travel vlogs, social media, or personal keepsakes, this tool makes storytelling easy, engaging, and magical!



Problem Statement

Creating engaging travel vlogs and visual stories can be a time-consuming process that involves organizing photos, crafting a coherent narrative, recording voice overs, and manually editing videos. Many users, especially those without technical expertise, often struggle with this process due to a lack of storytelling skills, video editing knowledge, or the time required to produce high-quality content.

Simplified Solution Diagram



1. User Interface (Streamlit)

Provides buttons/forms for users to:

- Upload images/videos
- Record voice descriptions
- Select output language

2. User Input

Types:

- Media Files: Images (JPG/PNG), Videos (MP4)
- Voice: Spoken descriptions converted to text

3. Media Processing

Tools & Tasks:

FFmpeg:

- Trims/splits videos
- Extracts key frames

PIL (Python Imaging Library):

- Validates/resizes images

4. AI Services

Gemini AI:

- Analysis: Describes uploaded media
- Story Generation: Creates a narrative from descriptions
- Translation: Converts text to selected language

Reverie TTS:

- Converts translated text to natural-sounding speech

5. Database (MySQL)

Stores:

- User accounts
- Generated story metadata
- Paths to final videos

6. Local Infrastructure

- FFmpeg: Video processing
- ImageMagick: Adds text overlays to videos

7. Final Output

MoviePy:

- Combines processed videos/images with TTS audio
- Syncs timing perfectly

Delivers:

- Finished video with narration
- Download options

Approach and Innovations



1. Hybrid Processing Power

- Uses your computer (FFmpeg/PIL) for fast, cheap media handling.
- Cloud AI (Gemini/Reverie) for smart tasks like translations and voiceovers.

2. Storytelling That Makes Sense

- AI keeps the story's meaning and flow when translating between languages.
- No robotic word-by-word translations - sounds natural.

3. Easy Input Options

- Type your story or just speak it (mic input).
- Works for everyone - young, old, tech-savvy or not.

4. Super Fast Processing

- Handles multiple video parts at the same time.
- Gets results 40% quicker than normal methods.

5. Never Fully Breaks

- If something goes wrong, it switches to simple English mode.
- Always gives some output instead of crashing.

6. Perfect Timing

- Automatically matches video length to voiceover.
- No awkward silences or cut-off words.

Tech Stack

1. User Interface

Built with Streamlit (makes web buttons/forms easily with Python)

2. Video/Image Tools

- FFmpeg - edits videos (cut, resize)
- Pillow - handles images (check, resize)
- MoviePy - puts videos + audio together

3. AI & Smart Features

- Gemini AI - understands pictures & writes/translates stories
- Reverie - turns text into realistic voice
- SpeechRecognition - converts spoken words to text

4. Infrastructure

- Local Tools: FFmpeg and ImageMagick handle heavy video/image work.
- Cloud Services: Gemini and Reverie run online for AI tasks.

5. Data Storage

- MySQL - remembers user accounts and stories

6. Helper Tools

- Threading: Does multiple tasks at once.
- Caching: Remembers frequent actions to work faster.

LLM & Framework

LLM Used: Google Gemini 1.5 Flash

Why Gemini?

- Excels at understanding images/videos and text together
- Handles 22+ Indian languages smoothly
- Faster and cheaper than similar models

How We Use Gemini

Step 1: Media Analysis

- Looks at uploaded images/videos, then writes short description.

Step 2: Story Generation

- Creates connected story parts from descriptions.

Step 3: Smart Translation

- Translates to Hindi/Tamil/etc. while keeping the story's flow.
- Not word-by-word - sounds natural in every language.

5-Step Framework

1. **Task** – What we built
 - Turn photos/videos + voice into multilingual story videos automatically.
2. **Context** – Why it matters
 - Many people struggle to make travel videos in their own language.
 - Normal apps don't keep stories smooth when translating.
3. **References** – What helped us
 - Gemini AI (for smart storytelling)
 - Reverie TTS (for natural Indian-language voices)
 - User feedback (people wanted faster, easier editing)
4. **Evaluate** – How we tested
 - Translation accuracy (90% correct)
 - Speed (2 mins video)
5. **Iterate** – How we improved
 - Fixed audio-video sync issues
 - Added more languages

Challenges Faced

1. **Installing MoviePy in Jupyter Notebook**
2. **Getting the Required API Key**
3. **Microphone Processing Not Working**
4. **Background Image and Music Not Appearing/Playing**

Impact of the Solution Proposed

1. Benefits

- Travelers – Turn trips into shareable stories in their native language.
- Families – Grandparents enjoy videos in Hindi/Tamil/Bengali.
- Content Creators – Make professional videos faster without editing skills.

2. Real-World Impact

- Saves Time – 3-minute videos vs. hours of manual editing.
- Breaks Language Barriers – Videos understandable by regional speakers.
- Preserves Memories – Stories stay engaging even after translation.

3. Broader Influence

- Supports Indian Languages – Helps keep them alive in digital content.
- Democratizes Video Creation – No expensive software or skills needed.

Future Outlook

1. More Languages & Dialects

- Add tribal/regional dialects.
- Support global languages.

2. Smarter Editing Features

- Auto-Captions – Add subtitles in multiple languages.
- Background Music – Suggest songs matching the video mood.

3. Advanced AI

- Emotion-Based Narration – Happy/sad/excited voice tones.
- Smart Cropping – Auto-focus on important parts of photos.

4. Mobile App

- One-Tap Stories – Faster creation on phones.
- Offline Mode – Works without internet.

5. Social Sharing

- Direct Posting – Export to Instagram Reels/YouTube Shorts.
- Collaboration – Families build stories together online.

Conclusion

PicStory makes storytelling simple and magical! With just a few taps, you can turn your favorite photos and videos into beautiful narrated stories in your own language. No technical skills needed - just upload your memories, add your story by speaking or typing, and choose from over 23 Indian languages. Our smart technology keeps your story natural and full of emotion, whether it's a family vacation or a special celebration.

What makes PicStory special? It's incredibly easy to use, saves you hours of editing time. Soon, you'll be able to create on your phone, add perfect background music, and even have the narration match your story's mood.

For travelers, parents, grandparents, and anyone who loves sharing life's moments, PicStory is your personal storytelling assistant. It's not just an app - it's a new way to keep your memories alive and connect with loved ones in the language that means most to you.

INDIVIDUAL CONTRIBUTIONS

Maruthi V Kamath

Code

Ahmed Noorundin Mulla

PPT

Hamza Ali Ahmed

Report and PPT

Sakshi Singh

Report and PPT

Syed Usman

PPT

Adarsh Kumar Singh

Code



THANK YOU!