



Image Description Generator

SnapNarrate



Team

22071A6746

Hansini

22071A6748

Rachana

22071A6750

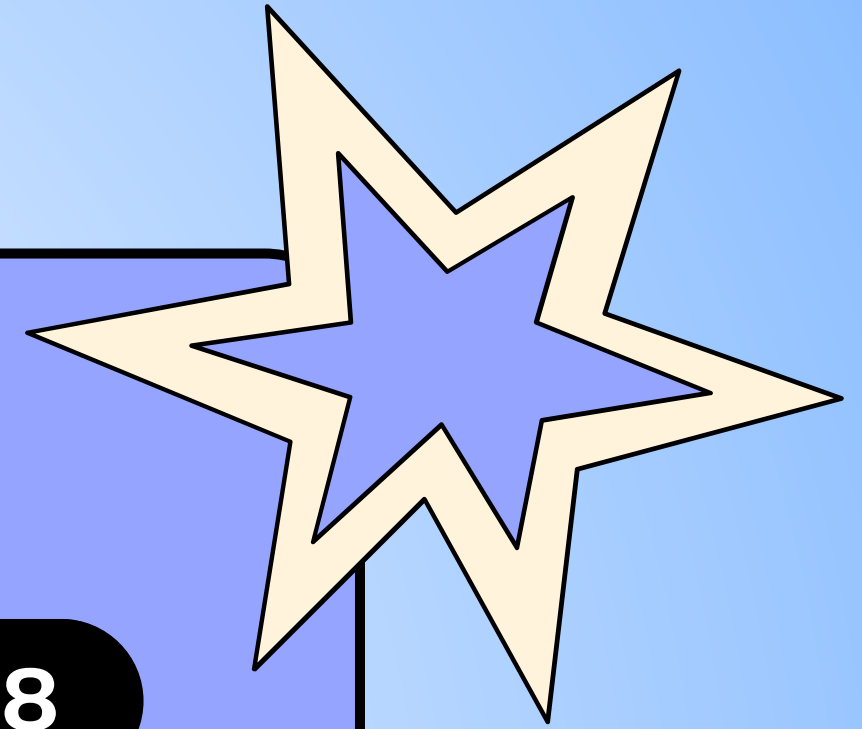
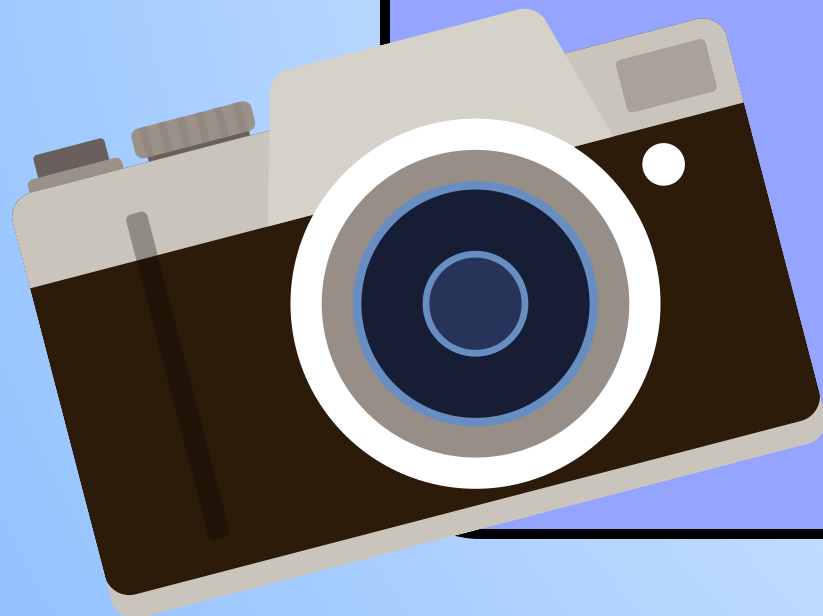
Radhika

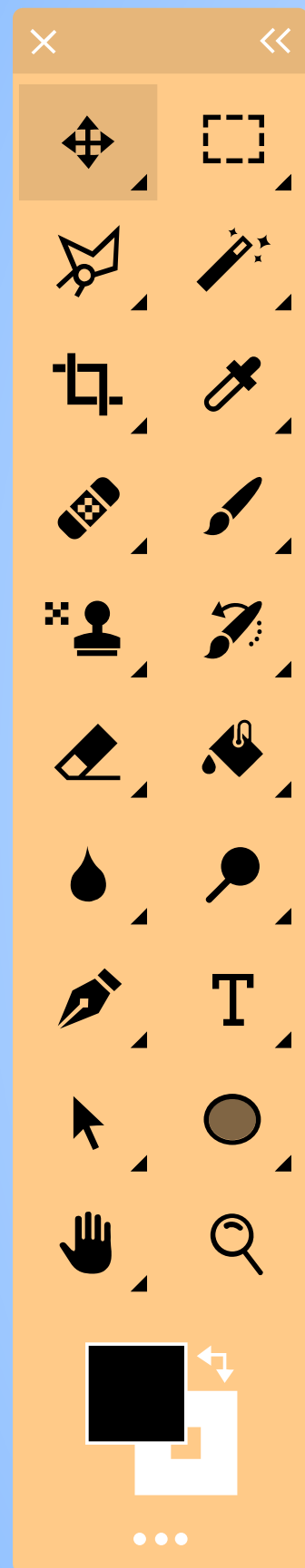
22071A6752

Sakshi

Guide

Mr. P. Veeranjaneeyulu





ABSTRACT

This project introduces an AI-powered system that transforms visual content into meaningful audio descriptions to support visually impaired users. Using the BLIP model for context-aware image captioning, it generates accurate and detailed descriptions which are then translated into English, Hindi, and Telugu using Google's Gemini API. These multilingual captions are converted into natural speech using Google Text-to-Speech (gTTS), offering a clear and accessible listening experience. The system supports multiple image input methods—upload, URL, or webcam—and runs efficiently on both local setups and cloud platforms like Google Colab. By integrating computer vision, language processing, and speech synthesis, the tool promotes digital inclusivity and empowers users to independently interpret complex visual scenes.



UPLOAD

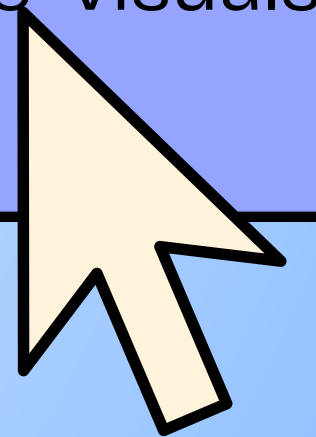
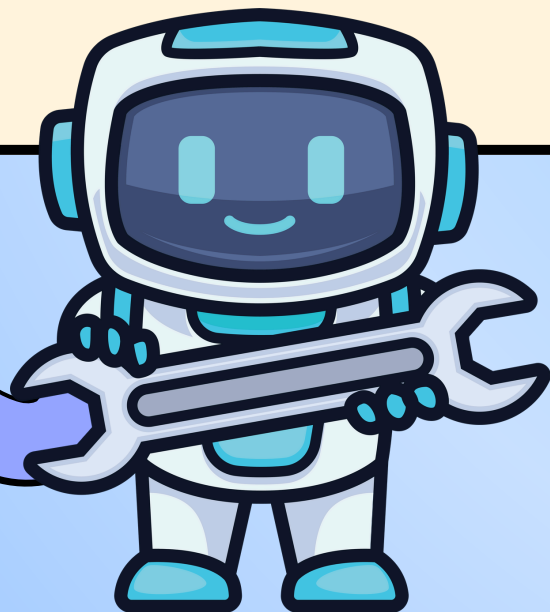


INTRODUCTION



Problem Statement

- Most image captioning tools generate basic, generic descriptions lacking depth and context.
- There's a growing need for detailed, expressive, and multilingual image understanding.
- Users from fields like education, accessibility, content creation, and language learning require tools that translate visuals into meaningful narratives.

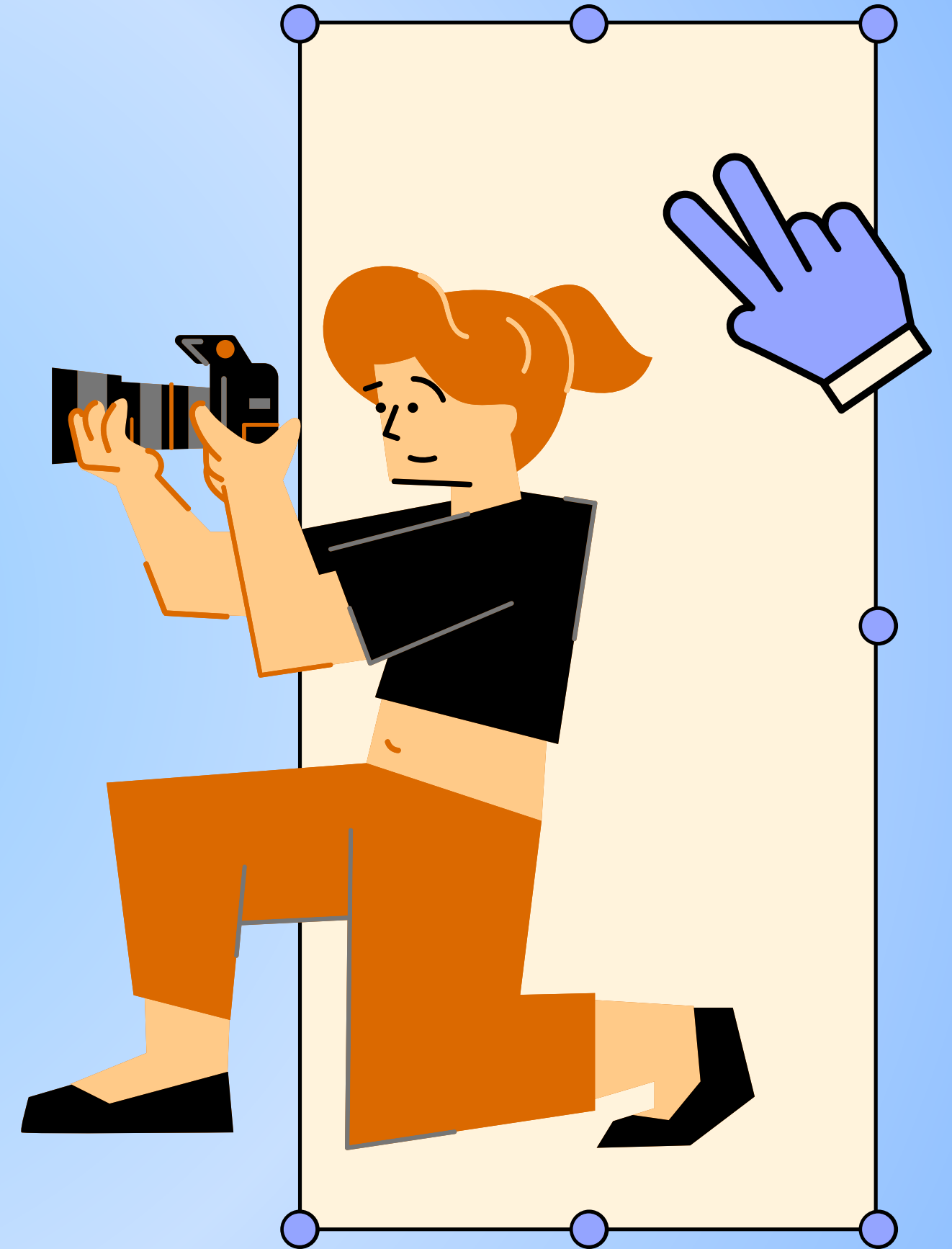


Objectives

Generate detailed, context-aware captions from images.

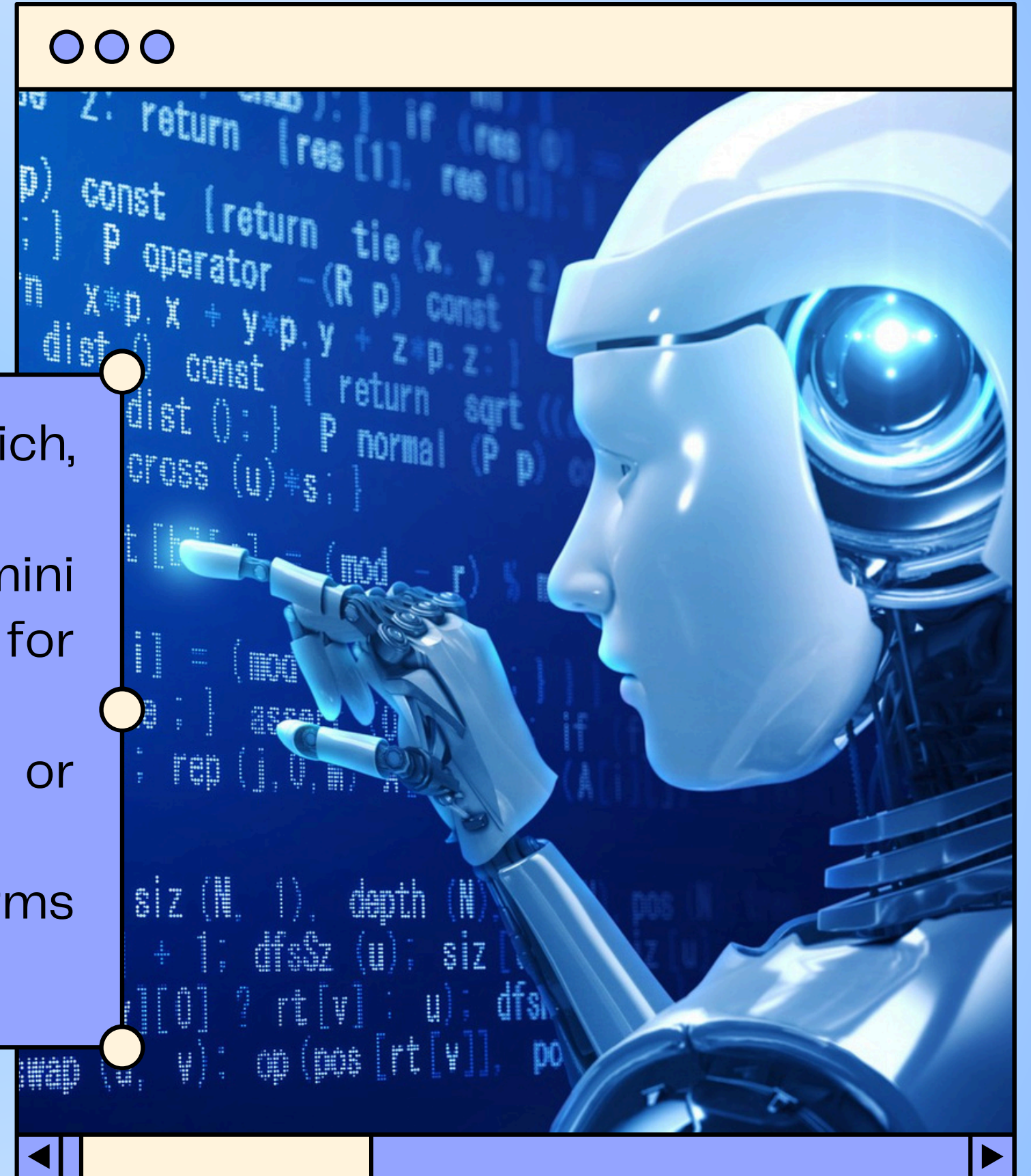
Provide multilingual and natural-sounding audio output.
Design a modular, scalable system for broad usability.

Enable use cases in accessibility, education, content generation, and language learning.



Project Scope

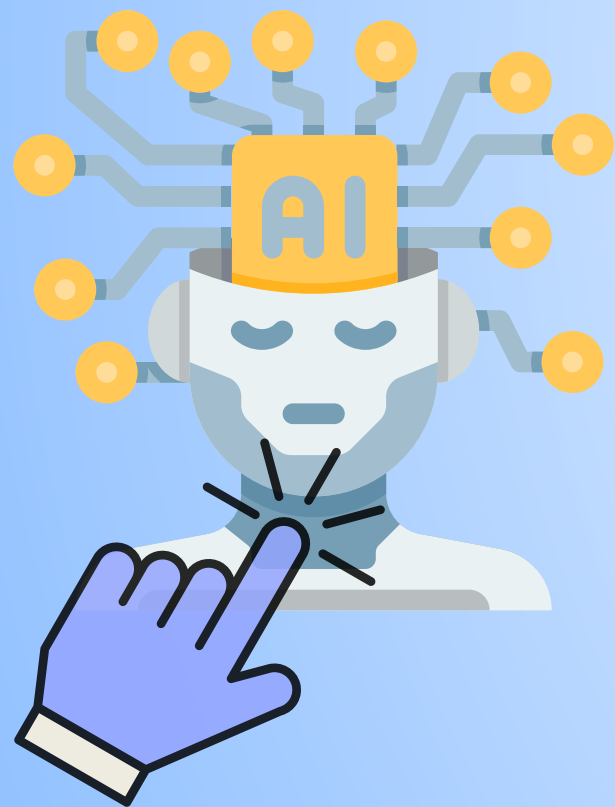
- * Build a system that converts images into rich, multilingual audio descriptions.
- * Use BLIP for image captioning, Google Gemini for enhancement and translation, and gTTS for speech output.
- * Support image input via upload, URL, or webcam.
- * Run on both local systems and cloud platforms like Google Colab.



Literature Survey

Current Systems

- Early systems like CNN-LSTM generated short, surface-level captions (e.g., "a cat on a bed").
- Lacked context, multilingual support, personalization, and audio output.
- Newer models like Transformer-based encoders and SmallCap improved relevance but still fall short in real-time and user-adaptive features.



A. Deep Learning Architectures

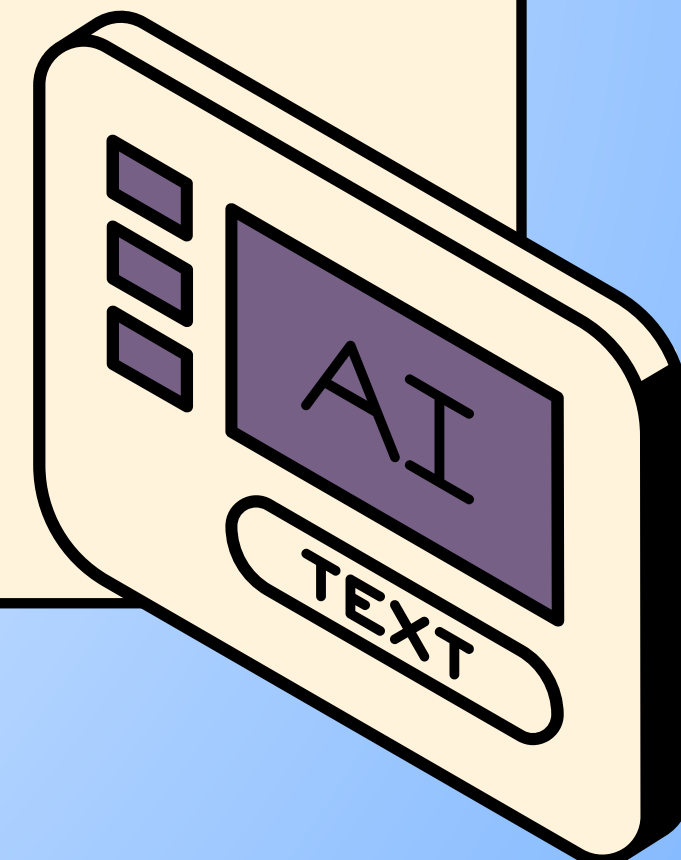
- CNNs + LSTMs enabled early captioning.
- Attention mechanisms improved contextual accuracy.
- Multimodal fusion & reinforcement learning increased adaptability.

B. Salesforce BLIP

- Uses Vision Transformers and prompt-based learning.
- Supports real-time, context-aware, and multilingual image descriptions.
- Outputs rich narratives (e.g., "a woman nervously pacing near a hospital bed").

C. Gemini & gTTS Integration

- Gemini translates and enhances captions into regional languages.
- gTTS converts them into natural-sounding speech.
- Forms a complete pipeline: Image → Text → Audio.



D. Limitations of Existing Systems

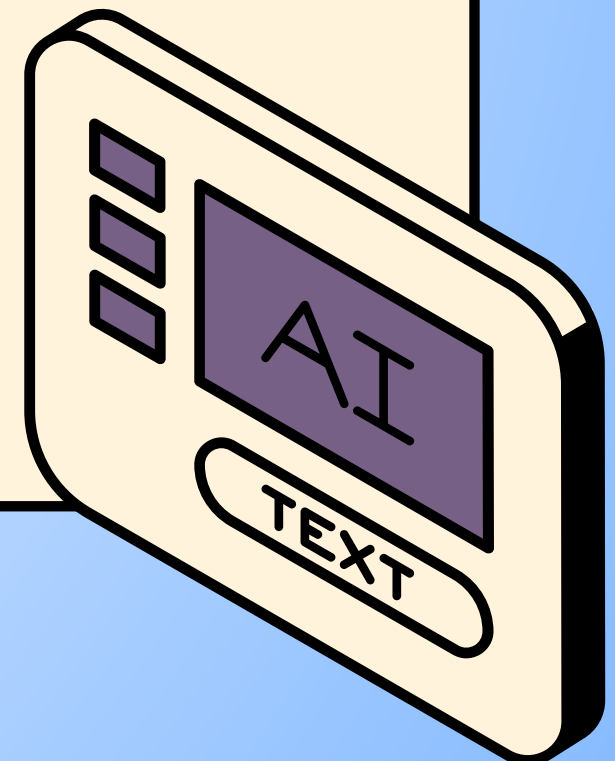
- Lacks interaction-level context, personalization, real-time input, and TTS integration.

E. Applications & Impact

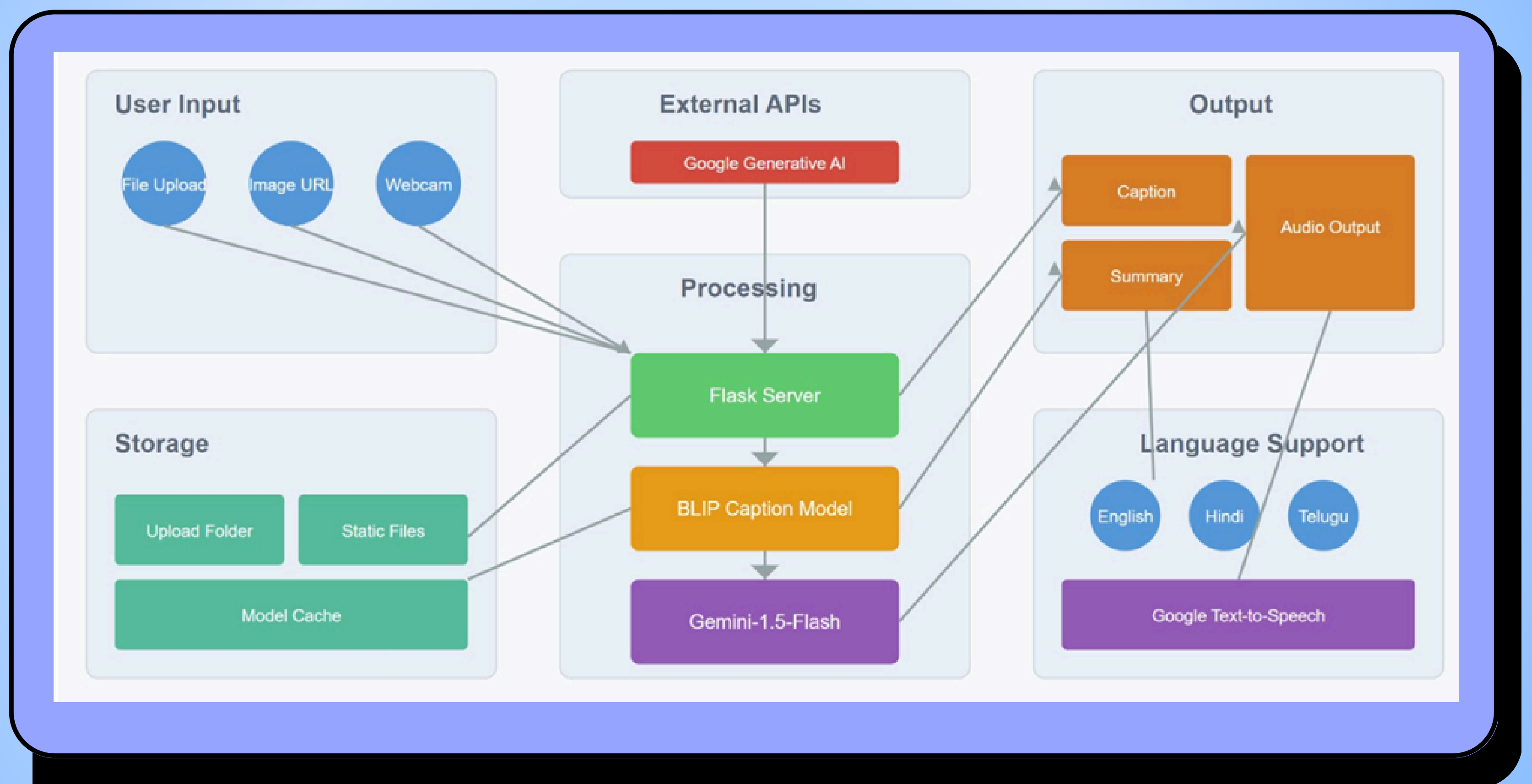
- Accessibility for the visually impaired.
- E-commerce: Better product description & SEO.
- Education: Helps blind students understand images.
- Social Media, Healthcare & Surveillance: Real-time, descriptive feedback.

F. Recent Trends (2022–2024)

- BLIP-2, SmallCap, Diffusion models, Tree-based outputs.
- Shift from captions to narrative, customizable descriptions.



Architecture Diagram



Proposed System

Investigation Phase

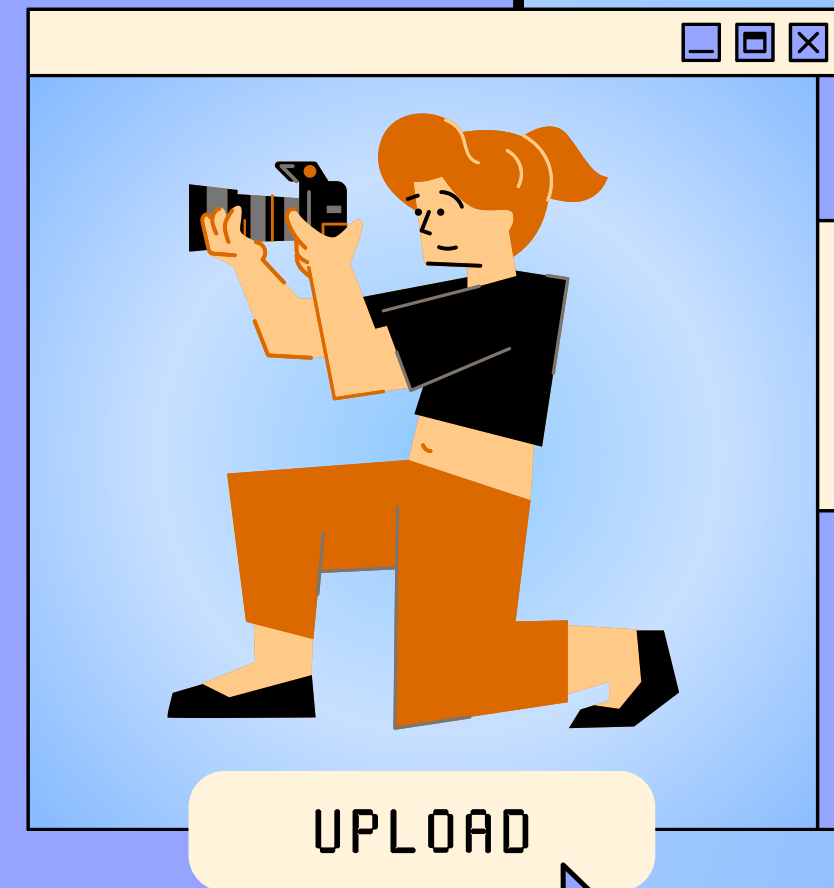
- Goal: Provide detailed image descriptions for visually impaired users with multilingual support and real-time processing.
- Image Input: Upload, live capture, or URL input.
- Preprocessing: Standardize images and extract features using CNN (ResNet).
- Use BLIP model for captioning.
- Translate captions to Hindi and Telugu via Google Gemini API.

Processing Phase

- Generate detailed image summaries using advanced language models.
- Convert text to speech with gTTS in multiple languages.
- User-friendly interface with interactive prompts and audio playback.
- Privacy ensured by local processing and no permanent image storage.

Maintenance Phase

- Regular updates for model improvements and language additions.
- Monitor performance via error logs and user feedback.
- Scalable system with resource management and cloud integration for translation and speech synthesis.



Conclusion & Future Scope



Our system makes it possible to **understand images through sound** by generating rich, **detailed descriptions** and converting them into speech. It opens up new ways for visually impaired users to explore visual content confidently, in the **language they're most comfortable** with.



In the future, the system can be improved to work


- with videos
- add emotion-based speech
- run on mobile phones


This will make it more helpful and easier for more people to use in everyday life.

Outcomes



Upload an image and get AI-generated captions, summaries, and audio narration

 Upload Image

 Paste URL

 Live Webcam


English



Choose an image from your device



Drag and drop an image or click to browse

 Generate Narration

Upload Image

Paste URL

Live Webcam


English

Choose an image from your device

Drag and drop an image or click to browse

Generate Narration

Your Image



AI Generated Content

Caption

a river with a few green plants in it

Summary

The image shows a tranquil river scene at sunset. A calm river meanders through a flat landscape of tall grasses. The river reflects the colorful sunset sky, which is a mix of oranges, pinks, and purples with dark clouds. A line of trees is visible in the distance along the horizon.

Audio Narration

0:00 / 0:22

Uploading
Image

Upload Image

Paste URL

Live Webcam

Hindi

Enter image URL

xcart.com/image/850/1000/kpcy5jk0/poster/h/c/w/large-village-poster-scenery-scenrym-68-original-imag3m8vrkdztzva.jpeg?q=20&crop=false

Image successfully uploaded and ready for processing

Generate Narration

Your Image



AI Generated Content

Caption

झील, नाव और पेड़ों की एक पेंटिंग

Summary

यहाँ दृश्य का वर्णन है: यह चित्र एक शांत झील के दृश्य को दर्शाता है, जिसमें एक छोटा सा, गुलाबी छत वाला पत्थर का घर एक छोटे से घाट के पास स्थित है। पानी शांत है और उसमें तीन हंस तैर रहे हैं। झील के किनारे पर, लाल पत्तियों वाले पेड़ और हरियाली दिखाई दे रही है। क्षितिज पर सूर्यास्त होता हुआ दिखाई दे रहा है, जिससे आसमान नारंगी और पीले रंगों से जगमगा रहा है। चित्र में कुछ पक्षी भी हैं, जो आसमान में उड़ रहे हैं।

Audio Narration

0:00 / 0:37

Image URL

 Your Image



 AI Generated Content


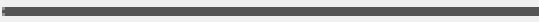


 Caption

జెల్లీ మరియు లైటు ఉన్న గది

 Summary

నేను దీన్ని సహాయపడతాను. ఇక్కడ ఇచ్చిన చిత్రంలో ఒక గోడపై బిగించబడిన ఏసీ యూనిట్ కనిపిస్తుంది. ఏసీ యూనిట్ తెల్లగా ఉంటుంది మరియు గోడ ఆకుపచ్చ రంగులో ఉంటుంది. గోడకు పైకప్పు కూడా కనిపిస్తుంది.

 Audio Narration

 0:00 / 0:16   

WebCam image input

Your Image



AI Generated Content

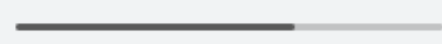


Caption

लकड़ी की मेज पर कॉफ़ी के दाने और एक कप कॉफ़ी की स्टॉक तस्वीरें

Summary

एक कप एस्प्रेसो भाप से उभर रहा है, जिसे एक छोटे से चम्मच से सजाया गया है, जो कॉफ़ी की फली के ढेर पर बैठा है, जो बर्लेप बैग से निकलता है। यह सब एक लकड़ी की मेज पर है।

Audio Narration

▶ 0:00 / 0:15   

AI Generated Content


Caption

coffee beans and a cup of coffee on a wooden table stock photos

Summary

Here's a description of the image: A white coffee cup filled with espresso sits on a small white saucer. Steam rises from the coffee. The cup and saucer rest on a scattering of dark brown coffee beans. A burlap sack filled with more coffee beans is partially visible in the background. The entire scene is set on a rustic wooden surface with the wood grain clearly visible. A silver spoon rests on the saucer next to the cup. A few additional coffee beans are scattered on the wooden surface around the cup and saucer.

Audio Narration

▶ 0:00 / 0:36   

AI Generated Content

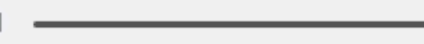

Caption

చెక్క బల్ల మీద కాఫీ గింజలు మరియు ఒక కప్పు కాఫీ స్టాక్ ఫోటోలు

Summary

అంధుల కోసం ఒక చిత్రం యొక్క సంక్షిప్త వివరణ ఇదిగో: ఒక చెక్క ఉపరితలంపై ఉన్న ఒక బుర్లాప్ సంచి నుండి కాఫీ గింజలు చిందిపోతున్నాయి. ఒక తెల్లటి కప్పులో కాఫీ ఉంది, దాని నుంచి ఆవిరి వస్తుంది. ఒక చిన్న చెంచా మరియు ఒక చిన్న పలక కప్పుకు అనుబంధించబడ్డాయి.

Audio Narration

▶ 0:00 / 0:21   

Output generated in different languages with same image

THANK
YOU!