IKSIKA.i Discover Life in Every Frame

Project By: Utsav Soni

- INDEX:

1. Introduction:

According to an estimate by the World Health Organization (WHO), 285 million people worldwide experience visual impairments, with approximately 39 million being legally blind. Recent statistics show that about 70% of visually impaired individuals in developed countries use smart devices, such as mobile phones and tablets, to improve accessibility and communication in daily life. The widespread usage of these devices in the visually impaired community is driven by the availability of features like screen readers, voice assistants, and high-contrast displays, which enable more effective interaction with technology.

Iksika/Ikshika, inspired by the Sanskrit word for "a glance" or "look," encapsulates the vision of this powerful AI application. Leveraging the capabilities of Meta's Llama 3.2 90B Vision Preview model and Groq's LPU Inference Engine, Iksika operates on Streamlit to analyze images captured in real time through a camera or uploaded from a device. In just moments, the app provides detailed descriptions of surroundings, translating its output into over 130 languages with text-to-speech functionality for immediate text and audio descriptions in the selected language.

Built with a focus on accessibility, Iksika aims to support visually impaired individuals by creating a more accessible world. Its intuitive UI allows for easy adaptation and interaction, empowering users to explore and understand their environment with just a device's camera and speaker. Beyond accessibility, Iksika offers versatile applications across social and business sectors, including tourism, translation, education, personalized assistance, and more.

2. Project Objectives:

"Iksika.i" aims to revolutionize the way visually impaired individuals interact with their environments by leveraging advanced vision AI technology. The project is dedicated to delivering precise, real-time feedback about surroundings, thereby enhancing navigation and situational awareness. A core objective is to facilitate communication across cultures by providing seamless multi-language translation, ensuring accessibility for a diverse user base.

The application is engineered with a scalable and efficient backend capable of processing high-resolution images swiftly, allowing users to either capture new images or upload existing ones for analysis. To further enrich user experience, features such as audio replay and regeneration will empower users to revisit descriptions at their convenience.

3. Innovation, Creativity, Business relevance and Potential impact:

"Iksika.i" showcases innovation through its unique live camera capture feature, a capability not widely explored by many developers in the assistive technology space. Alongside this, the app supports image uploads from local storage, providing users with flexibility in how they interact with the application. Unlike typical applications that focus solely on generating English-language outputs, "Iksika.i" integrates a live translation feature using the deep-translator library, enabling real-time conversion of text into any selected language. This is complemented by a text-to-speech function powered by the gTTS library, enhancing accessibility for visually impaired users who often rely on screen readers.

The user interface is meticulously designed to cater to visually impaired individuals, ensuring easy navigation through sound cues associated with each button, thereby facilitating a seamless user experience. Additionally, the app is not only tailored for visually impaired users but also offers valuable functionality for the general population in everyday contexts. By fine-tuning the Llama 3.2 model, powered by Groq, for specific industry applications, "Iksika.i" aims to scale its capabilities and reach an expanding user base.

The social impact of this app is significant, as it is freely accessible on the Streamlit platform, ensuring that anyone can benefit from its features without financial barriers. This commitment to inclusivity not only supports individuals with visual impairments but also promotes broader societal engagement and awareness, potentially transforming how technology is perceived and utilized across diverse communities.

4. Features and Functionalities of Iksika:

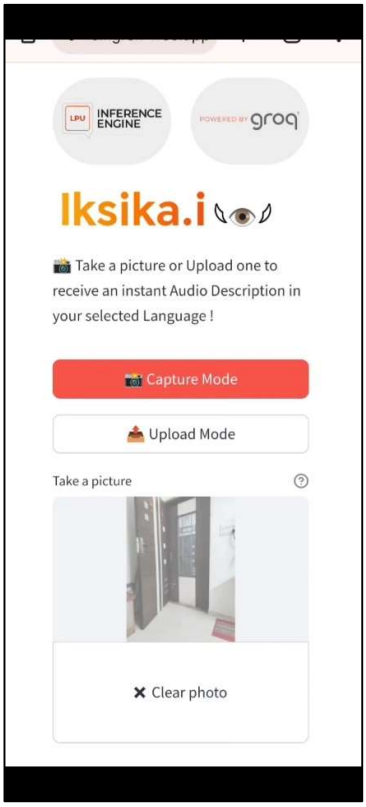
- Real-Time Image Analysis: Capture live images or upload them from your device, and instantly get a detailed description of your surroundings in your language.
- Groq Powered LLaMA 3.2 Vision Model: Utilizes Meta’s LLaMA 3.2-90B Vision Preview model, optimized with Groq's inference engine for high-performance image recognition and analysis.
- Multi-Language Support: Provides descriptions in over 50 languages, making it accessible to users globally with customizable language selection.
- Instant Talkback: Generates text-to-speech output in the selected language, enabling users to hear the description for hands-free interaction.
- Regenerate Description: Offers a "Regenerate Description" button for obtaining an updated or alternative description with a single click, ensuring accurate insights.
- Replay Audio: Features a "Replay Audio" button to easily listen to the description again, providing an effortless way to revisit information.
- Traceable Insights with LangSmith: Integrates LangSmith for traceable insights, offering transparent tracking and logging for analysis accuracy.
- Intuitive UI for Accessibility: Designed with a user-friendly interface that’s easy to navigate and accessible to visually impaired users.
- Modular Code Structure: Organized in modules to support easy scalability, customization, and integration of additional features.
- Open-Source Library Integration: Uses a suite of powerful libraries, including Groq, Langsmith Streamlit, gTTS, Logger, Pillow and Deep Translator, for seamless functionality.
- Designed for Diverse Use Cases: Adaptable for various applications, from assisting visually impaired users to supporting tourism, education, and business needs.

5. Results and Evaluation:

Run Count (7D)	Error Rate (7D)	% Streaming (7D)	Total Tokens (7D)	Total Cost (7D)	P50 Latency (7D)
167	0%	0%	[redacted]		3.17s

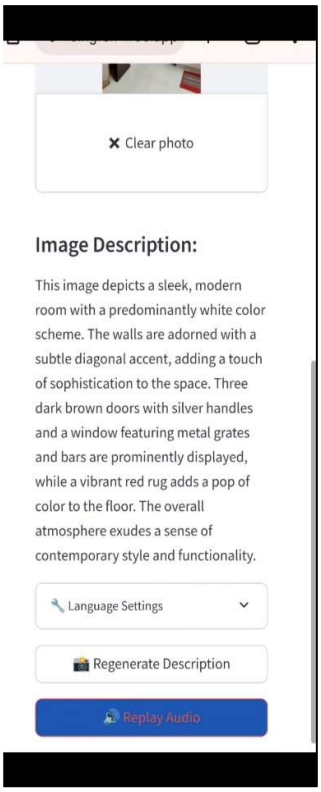
- Iksika demonstrates exceptional performance metrics with a remarkable 0% error rate across 167 test runs, showcasing its reliability in real-world applications. The application consistently delivers results within an impressive P50 latency of 3.17s, making it highly responsive for real-time assistance. During testing, the system maintained a streamlined 0% streaming overhead while efficiently processing image analysis and language translation across 130+ languages. The scalable architecture, powered by Groq's LPU Inference Engine and Meta's Llama 3.2 90B Vision Preview model, handles complex image recognition tasks with high accuracy and minimal resource utilization. Most notably, the system's rapid response time of approximately 10 seconds for complete analysis, translation, and audio generation demonstrates its practical viability for real-world deployment.

6. Considerations during UI Design:



- Added Capture photo and Clear photo button at the bottom of the mobile screen, so visually impaired person can easily click the button.
- Each and every button on the app has sound, as an when user click on the button he'll be notified by the sound.
- Capture mode and Upload mode button are almost centre aligned with the mobile screen so user can easily found that button.
- And after description generated, once if user do one scroll down, then next set of button will also come at the bottom of the screen again which is also added sound in it.

(For Promotional reasons
LPU inference Engine and powered by Groq Icon kept on top which sprinkles as eye kind of effect on App, for general audience specifically.)



- Demo video:
  https://drive.google.com/file/d/1k4gXkMPdp_3y2ceDt5ormyv9JzvhNVxJ/view?usp=sharing

7. Code Structure:

```
├── iksika.py              # Main entry point for the Streamlit app
├── .env                   # Environment variables file (keep it secure)
├── requirements.txt   # Project dependencies designed in such way to support scaleability
├── readme.md          # Short and Crisp readme file
├── logs/                  # Log files directory
|   └── app.log            # Log file for application activity
├── assets/                # Images, icons, and audio files
|   └── mark1.svg        # groq logo
|   └── mark2.svg        # groq logo
|   └── camera.mp3      # mp3 on button clicks
|   └── button1.mp3     # mp3 on button clicks
|   └── button2.mp3     # mp3 on button clicks
|   └── button3.mp3     # mp3 on button clicks
├── static/                # directory for css
|   └── style.css          # CSS file for webApp
├── src/                   # Source code directory
|   └── __init__.py      # Makes the directory a package
|   └── tts_manager.py   # Handles text-to-speech operations
|   └── utils.py             # Utility functions (e.g., image handling)
|   └── image_to_text.py    # Handles the image-to-text logic
|   └── logger.py        # Central logging setup
├── tests/                 # Test cases
|   └── test_utils.py   # Tests for utility functions
├── tests/                 # Test cases
```

8. Conclusion and Future Scope:

Iksika.i offers a uniquely adaptive approach to image-based assistance, bringing advanced AI-driven visual insights to visually impaired individuals while remaining versatile for broader applications. Designed with accessibility and global inclusivity at its core, the app facilitates real-time environmental descriptions, supports over 130 languages, and delivers multi-sensory feedback through text-to-speech. Iksika.i extends its potential into various fields, including tourism, language learning, art therapy, wildlife exploration, and even emergency aid, highlighting its extensive adaptability for daily, professional, and recreational use. With its future capabilities, Iksika.i stands as a transformative tool, bridging gaps in accessibility, knowledge, and experience across diverse user groups, fostering independence and empowerment in meaningful ways.

Future Scope:

Tourism: Use images of famous place or monument or difficult-to-reach places to create immersive experiences for people who visit physically or upload images, describing unique details of place.

Personalized Workout Assistant: Recognize gym equipment and provide exercise guidance, especially in local languages, helping beginners use equipment safely.

Wildlife Exploration for Hikers: Assist hikers or explorers by identifying flora, fauna, or even animal tracks and offering real-time insights in a chosen language.

Educational Aid for Visual Learning: Enable students to photograph academic materials or visuals, offering detailed, language-specific explanations that support remote or self-guided learning.

Language Learning Through Visual Cues: For language learners, offer word translations and cultural context by interpreting everyday objects they encounter.

Food Identification and Explanation in Foreign Cuisine: For travellers, identify dishes and ingredients in a restaurant menu by image capture, providing details on preparation or origin.

Recipe Book Companion: Recognize ingredients in a user's kitchen to suggest recipes or nutritional insights, with explanations in their preferred language.

Personalized Art Therapy: Enable users in art therapy sessions to capture and receive feedback on their artwork, providing descriptions that foster self-reflection or therapeutic insights, or get description on famous art work by just click picture of it.

Home Décor Suggestions: By scanning an existing room, the app can suggest décor items or furniture arrangements and describe them to match the user's language preferences.

Interior Mapping for the Visually Impaired: In large indoor areas like offices or malls, provide detailed layouts by interpreting directional signs and guiding users through complex routes.

Pet Identification and Care Tips: Capture pet images to provide species information, potential health tips, and recommended care practices tailored to different pet types.

Gardening Assistant for Urban Farmers: Recognize plant health by image and offer insights on care, soil recommendations, and sunlight needs, supporting urban agriculture.

Waste Management and Recycling Assistance: Help users identify and sort waste materials by capturing images of items, **promoting sustainable practices** by categorizing recyclables and compostables.

Emergency First Aid Assistant: Recognize injuries or medical situations from images and provide step-by-step first aid instructions, potentially saving lives in critical situations.

9. Contact Information:

Name: Utsav Soni

From: Gujarat, India.

Mail id: soniutsav22@gmail.com

Linked In: www.linkedin.com/in/utsav-soni-9067b31b3

----------| THANK YOU |----------