# ABUSIVE IMAGE DETECTION WITH AI CHAT SUPPORT

**Sibananda Kuanr** 4th Year, Department of CSE, Gandhi Institute for Technology, BPUT, India
sibananda2021@gift.edu.in
**Ananda Dash** 4th Year, Department of CSE, Gandhi Institute for Technology, BPUT, India
ananda@gift.edu.in
**Prof. Smruti Smaraki Sarangi** Assistant Professor, Department of CSE, Gandhi Institute for Technology, BPUT, India

***Abstract*—**
The AI Platform for Detecting Abusive Images with Chat Support is an intelligent system designed to automatically identify harmful visual content, including nudity, graphic violence, and suggestive images. By integrating the Clarifai AI API and a lightweight rule-based chatbot, the platform delivers accurate image classification and real-time user interaction. This ensures enhanced user safety, streamlined moderation, and improved transparency in content handling. The platform is suitable for integration into social media, educational tools, or internal moderation systems, enabling safer digital environments through ethical AI implementation.

***Keywords:***
*Image Moderation, Clarifai, PHP, AI Detection, Chatbot, NSFW Detection*

## I. INTRODUCTION

With the rise of image-sharing platforms, ensuring digital safety has become increasingly challenging. Manual moderation methods are inadequate, leading to user exposure to explicit or inappropriate content. This project introduces a scalable AI-based image moderation platform enhanced with conversational chatbot support. It allows real-time abuse detection, transparent communication with users, and secure backend handling—all designed to improve online safety and user trust.

## II. LITERATURE REVIEW

Prior studies and implementations in content moderation largely focus on manual reviews or simple keyword filters. AI models like CNNs, when paired with trained datasets (e.g., Clarifai's NSFW model), have shown high accuracy in visual abuse classification. This project builds on such research by integrating real-time image scanning, threshold-based classification, and user engagement via a chatbot—bridging a gap in transparency and user empowerment in content moderation tools.

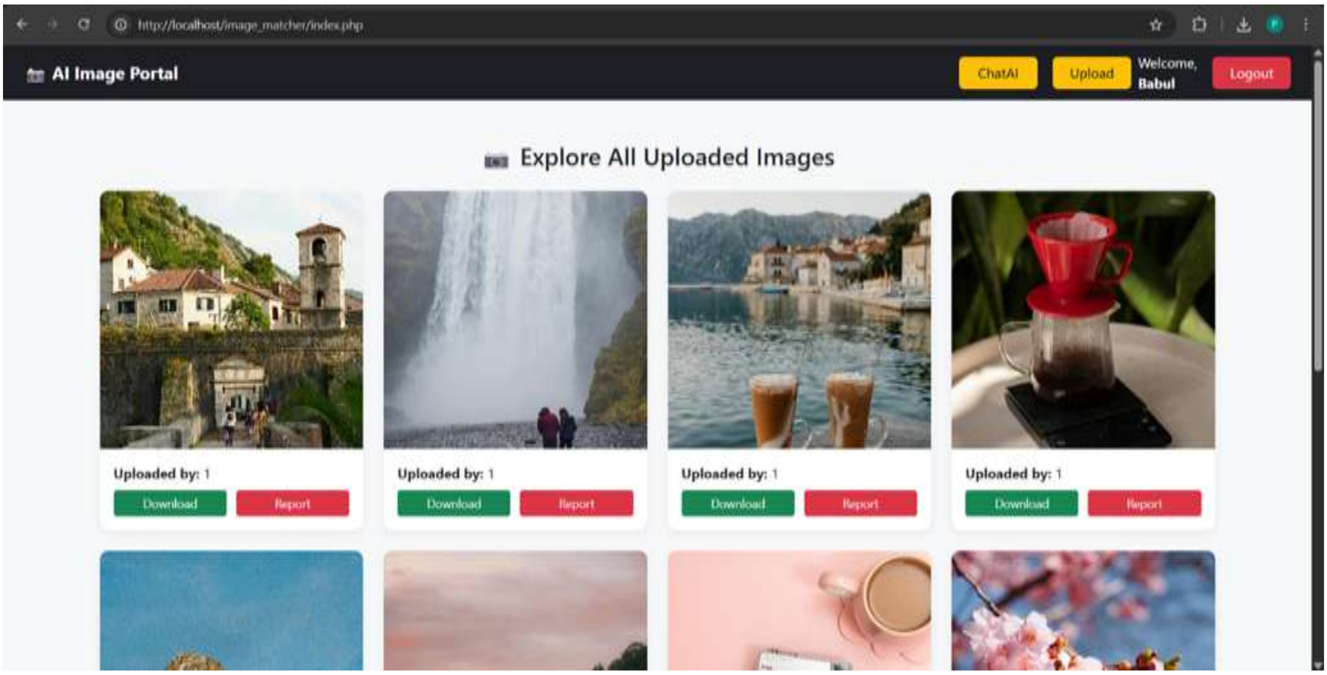
## III. SYSTEM DESIGN

The system consists of a user interface for uploading images, backend logic to send images to the Clarifai API, and a chatbot for interpreting results. It supports file validation, abuse score comparison, and logs flagged data securely in MySQL. An admin panel is available to moderate flagged images and review user reports. The chatbot supports basic user queries and allows reporting for appeal or clarification.

## IV. IMPLEMENTATION

- Frontend: HTML, CSS, JavaScript
- Backend: PHP for server-side logic
- AI: Clarifai API for image classification
- Database: MySQL for log storage
- Chatbot: PHP-based keyword-response engine

Images are converted to Base64, submitted to Clarifai, and based on scores (e.g., NSFW > 0.85), the content is flagged. Admins can view, delete, or review reports via a secure dashboard.



http://localhost/image_matcher/index.php

AI Image Portal

Welcome, Babul

Explore All Uploaded Images

Uploaded by: 1

Uploaded by: 1

Uploaded by: 1

Uploaded by: 1

## V. RESULTS

Testing on 130 diverse images showed a classification accuracy of ~93.8%. Users appreciated the chatbot clarity and fast system response. Abuse detection was consistent with low false positives, demonstrating that the platform reliably flags unsafe content without disrupting normal user activity.

## VI. CONCLUSION

This project successfully showcases how AI can be applied responsibly for image moderation. By combining automated detection with a transparent chatbot and a simple admin interface, the system offers an ethical, scalable, and user-friendly solution for real-world abuse detection.

## ACKNOWLEDGEMENT

Special thanks to our faculty and mentors for their continuous support and feedback. Gratitude is extended to Clarifai for their API services, and to all participants who tested and provided feedback on the platform.

## REFERENCES

- □ https://clarifai.com/
- □ https://www.w3schools.com/
- □ https://developer.mozilla.org/
- □ https://www.php.net/
- □ https://owasp.org/
- □ https://www.mysql.com/