IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

"Unveiling The Truth: Exploring AI Solutions To Identify Real Vs Synthetic Images"

[1]Mr. Animesh Sanjay Timande, [2]Mr. Gorakshan Bapurao Ingole, [3]Mr. Rutvik Vilasrao Behare. [4]Mr.Purvesh Purushottam Dabhade, [5]Prof. Snehal V. Raut,

[1,2,3,4] Student, [5] Prof

1,2,3,4 Student, Dr. Rajendra Gode Institute of Technology and Research, Amravati, IN,

5Guide, Prof Dr. Rajendra Gode Institute of Technology and Research, Amravati, IN

Abstract: In an era where synthetic media and deepfakes are becoming increasingly sophisticated, ensuring the authenticity of digital content has emerged as a critical challenge. This project explores an innovative AI-driven solution to distinguish real images from synthetic ones by utilizing the capabilities of Large Language Models (LLMs), specifically through Gemini technology. Unlike traditional methods that rely solely on pixel-level analysis, our approach leverages the multimodal understanding of LLMs to interpret visual data with contextual depth and semantic reasoning. The system is deployed via a web-based platform, supported by cloud services like Supabase for storage and authentication. Through this work, we aim to enhance digital trust by providing a scalable, intelligent tool for detecting manipulated visual content, contributing to the broader field of AI-based media forensics.

Index Terms **-** Real vs Fake Image Classification, Image Authenticity Verification, Synthetic Media Analysis, AI-Based Media Forensics, Digital Image Integrity.

I. INTRODUCTION

1.1 Preface

In recent years, the proliferation of artificial intelligence, especially generative models like GANs (Generative Adversarial Networks), has brought about remarkable innovations in image synthesis. These AI-generated images have reached such a level of photorealism that distinguishing between real and synthetic visuals has become increasingly challenging, even for trained human eyes. While these advancements offer exciting applications across entertainment, design, and education, they also pose significant risks in terms of misinformation, identity fraud, and deepfake-based deception.

This project, titled "Unveiling the Truth: Exploring AI Solutions to Identify Real vs Synthetic Images", is a step towards addressing the growing need to develop reliable AI-powered tools capable of discerning the authenticity of images. It aims to explore cutting-edge deep learning techniques and computer vision algorithms that can effectively distinguish between real and AI-generated images.

The work presented in this thesis encapsulates the learning, development, and experimentation carried out to understand the depth of the problem and propose a feasible solution. The project reflects not only an academic pursuit but also a contribution to the broader societal need for digital truth verification in the age of artificial media.

1.2 Motivation

The motivation behind this project arises from the growing influence of generative AI and the concerning rise of synthetic media. Today, tools like DALL·E, Midjourney, and Stable Diffusion enable users to generate highly realistic images within seconds. While these tools have democratized creativity, they have also led to a surge in manipulated content online, some of which is indistinguishable from reality.

From political deepfakes to synthetic celebrity images and misleading news visuals, fake media has the power to misinform the public and erode trust in digital content. The implications are serious—ranging from threats to personal privacy to the potential manipulation of democratic processes.

This project is motivated by a desire to counteract these challenges by creating a system that utilizes AI to fight AI-generated deceptions. It aims to bridge the gap between AI creativity and AI accountability by enabling users, platforms, and organizations to verify image authenticity efficiently and accurately.

1.3 Problem Statement

With the exponential advancement in generative AI, distinguishing real images from synthetically generated ones has become a pressing challenge. Current detection methods are often limited by generalization issues, dataset biases, and an inability to keep up with the rapidly evolving quality of generated images.

The core problem this project addresses is:

"How can we build an effective AI-based system to accurately differentiate between real and AI- generated (synthetic) images, considering the increasing realism of generative models and the dynamic nature of visual datasets?"

This question encapsulates the technical, ethical, and societal dimensions of the problem. The challenge lies not only in building an accurate classification model but also in ensuring that it adapts to new styles of image generation and maintains robustness across different data sources.

1.4 Objectives

The primary objective of this project is to develop a machine learning model capable of identifying whether an image is real or AI-generated. This objective is supported by the following sub-goals:

- To study and analyze the characteristics of synthetic images generated by various AI models.
- To collect and preprocess datasets comprising both real and AI-generated images.
- To implement and compare different deep learning architectures (e.g., CNNs, ResNet, EfficientNet) for classification tasks.
- To evaluate the performance of the models using metrics like accuracy, precision, recall, and F1 score.
- To identify patterns or visual artifacts commonly found in synthetic images.
- To explore the use of explainable AI (XAI) techniques to interpret model decisions and improve transparency.
- To provide a prototype or demo system that can detect image authenticity in real time or batch processing.
- These objectives guide the project through its research, development, and testing phases and aim to contribute a reliable solution for real-world applications.

II. SCOPE

The scope of this project is centered on the development and evaluation of artificial intelligence (AI) techniques to distinguish between real and synthetic (deepfake or AI-generated) images. With the growing accessibility of powerful generative models such as GANs (Generative Adversarial Networks), it has become increasingly important to build robust detection mechanisms that can identify manipulated or fabricated visual content.

2.1 Dataset Collection and Preprocessing:

The project involves gathering diverse datasets containing both real and synthetic images from publicly available sources. The data is preprocessed to ensure consistency, enhance quality, and prepare it for model training and evaluation

2.2 Model Development Using Deep Learning:

AI models, particularly deep learning architectures like CNNs (Convolutional Neural Networks), are implemented using TensorFlow. These models are trained to extract subtle patterns and anomalies that distinguish synthetic images from authentic ones.

2.3 Performance Evaluation:

The models are evaluated using various performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The goal is to determine their effectiveness in identifying synthetic content across different scenarios and datasets.

2.4 Real-World Application and Deployment:

The project explores practical implementation through a web-based application that integrates AI models via backend services (LLM-gemin technology) and frontend technologies (HTML, CSS, JavaScript, React.js or Angular). Supabase is utilized for cloud storage and authentication.

2.5 Ethical and Social Considerations:

The project also examines the ethical implications of deepfake technology and the societal importance of reliable detection methods to combat misinformation, identity theft, and digital fraud.

III. LITERATURE SURVEY

The rapid advancement in AI technologies has led to the emergence of highly realistic synthetic images generated by modern generative models, particularly diffusion-based techniques. These images are so convincingly rendered that distinguishing them from authentic visual content has become increasingly difficult, presenting significant challenges to digital media verification and content integrity. In this chapter, we explore existing literature on synthetic image generation and methods for detection, particularly focusing on non-CNN-based approaches and the integration of newer AI technologies.

3.1 Generative models: A Brief Overview

Generative models such as diffusion models have revolutionized synthetic content creation. These models operate by reversing a noise diffusion process to gradually construct detailed and realistic images. Notable examples include:

- Stable Diffusion and DALL·E 2: Capable of generating high-quality, photorealistic images based on textual prompts, these models have pushed the boundaries of synthetic media creation.
- Unlike traditional models, diffusion approaches produce clearer images with fewer artifacts, making detection increasingly challenging.

The growing sophistication of such generative tools underscores the need for reliable detection systems that do not rely solely on conventional convolutional techniques.

3.2 Importance of Detection Systems

Accurate detection of synthetic images is essential in several real-world contexts:

- Preventing the spread of misinformation
- Maintaining trust in journalism, law enforcement, and surveillance systems
- Protecting individuals from digital impersonation or fraud

With traditional digital forensics unable to keep pace with current generative capabilities, researchers are shifting toward AI-powered solutions, including language model–based reasoning and hybrid detection techniques not dependent on CNNs or GANs.

3.3 Modern Detection Approaches without CNNs

While CNNs have historically been effective in image detection, modern systems are increasingly exploring alternative strategies, including:

Language Model-Based Reasoning (LLMs)

With the advent of LLMs (Large Language Models), such as those used in Gemini Technology, detection can be approached from a semantic and contextual analysis perspective. By leveraging multimodal understanding, LLMs can:

- Analyze associated metadata and captions
- Evaluate image context for inconsistencies
- Aid in detecting synthetic content without relying on pixel-based convolutional analysis

3.4 System Infrastructure & Technologies Used

Our detection platform is built on a lightweight yet powerful tech stack designed for efficiency, scalability, and ease of access:

- Frontend: Developed using standard web technologies (HTML, CSS, JavaScript) for seamless user interaction and cross-device compatibility.
- Backend: Utilizes Gemini Technology, incorporating advanced LLMs for inference and semantic reasoning to assist in detection workflows.
- Authentication and Storage: Handled through Supabase, a cloud-native backend that ensures secure user authentication and image storage.

3.5 Datasets for Detection

Although traditional CNN-based models rely on datasets such as CelebA-HQ or DFDC, our system leans on custom image collections enriched with metadata and generative context, allowing LLMs to infer potential inconsistencies or synthetic artifacts based on broader multimodal patterns. This helps mitigate dataset-specific biases and enhances generalization to newer synthetic content types.

IV. System Architecture

The proposed system is a web-based application designed to determine whether an image is real or synthetically generated using artificial intelligence. It uses a simplified, efficient, and scalable structure composed of three core layers: Frontend User Interface, Backend Processing System, and Machine Learning and Cloud Storage Layer. This modular design ensures smooth functionality, ease of maintenance, and adaptability to future enhancements or scale-ups.

4.1 Overview of the System Architecture

The architecture is developed using a lightweight approach, minimizing complexity while ensuring high responsiveness and real-time performance. Technologies are chosen to meet the requirements of accessibility, scalability, and secure data handling without relying on complex frontend or backend frameworks.

4.2 Frontend User Interface Layer

The Frontend User Interface (UI) serves as the user's primary interaction point with the system. It is developed using core web technologies—HTML, CSS, and JavaScript— with a focus on simplicity, responsiveness, and clarity. The goal of the frontend is to make it easy for users to upload images and receive immediate feedback about whether an image is AI-generated or real.

Key Features of the Frontend Include

- User Registration and Login: Users can create accounts and log in securely through a clean, form-based interface that communicates with Supabase for authentication.
- Image Upload Interface: Users can upload images either by selecting files from their device or using drag-and-drop functionality. The design ensures this process is quick and intuitive.
- Real-Time Feedback Display: As soon as the image is uploaded, the frontend displays whether the image is AI-generated or authentic, based on results returned from the backend.
- Responsive Web Design: The layout is mobile-friendly and responsive across different devices and screen sizes, ensuring consistent user experience on desktops, tablets, and smartphones.

- User Registration and Login: Users can create accounts and log in securely through a clean, form-based interface that communicates with Supabase for authentication.
- Image Upload Interface: Users can upload images either by selecting files from their device or using drag-and-drop functionality. The design ensures this process is quick and intuitive.
- Real-Time Feedback Display: As soon as the image is uploaded, the frontend displays whether the image is AI-generated or authentic, based on results returned from the backend.
- Responsive Web Design: The layout is mobile-friendly and responsive across different devices and screen sizes, ensuring consistent user experience on desktops, tablets, and smartphones. The absence of frameworks like React or Angular is compensated by using **pure JavaScript** for asynchronous functionality, such as handling image uploads and dynamically updating content without refreshing the page.

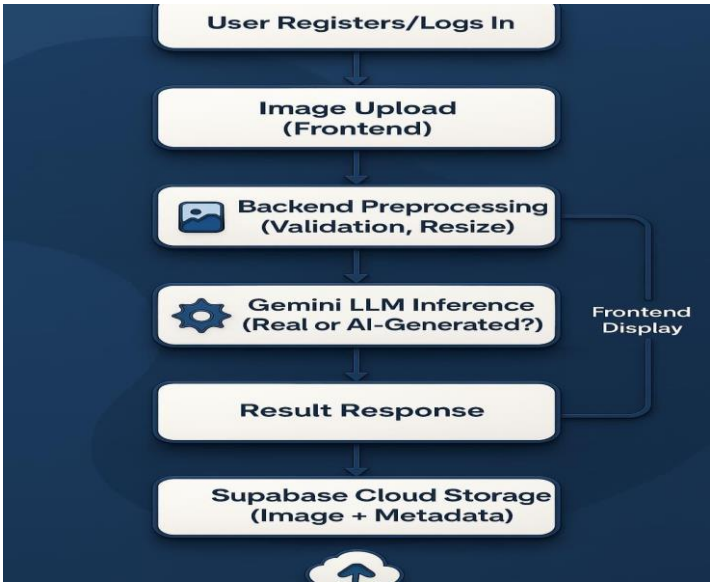

Fig 1. System Architecture Diagram

4.3 Backend Processing System Layer

The Backend Processing **System** is the core logic layer that connects the frontend to the intelligent inference system. Unlike traditional setups using CNNs or DNNs, this backend employs Gemini Technology integrated with Large Language Models (LLMs), capable of multimodal processing and understanding of image data.

Key Function of the Backend Include

- **API Endpoints**: The backend exposes RESTful APIs built with simple HTTP handling logic in either Node.js or Python. These endpoints receive image uploads, manage user sessions, and return classification results.
- **User Session Handling**: Through integration with Supabase, the backend securely manages user sessions using modern token-based authentication mechanisms such as JWT (JSON Web Token).
- **Image Routing and Processing**: The backend receives uploaded images from the frontend, forwards them to the Gemini-powered model for analysis, and returns the results back to the frontend.
- **Asynchronous Handling**: Since image analysis can take a short processing time, the backend handles tasks asynchronously to avoid blocking the server during inference.

Gemini's LLM is a cloud-accessible AI service that is capable of understanding visual input and determining if it carries markers of AI-generated content. This includes detecting irregularities or common traits associated with synthetic media.

4.4 Machine Learning and Cloud Storage Layer

This layer is built upon two foundational components:
Gemini Large Language Model(LLM)

- Gemini's LLM provides the system’s intelligence. It supports image-based inputs and is designed for multimodal analysis, meaning it can interpret visual data with contextual awareness.
- Rather than using manually trained CNN models, the system leverages Gemini's advanced understanding to

identify artifacts, lighting inconsistencies, unnatural patterns, or other signals commonly found in AI-generated images.

- Since Gemini operates as a service, it minimizes the need for local model training or heavy computation infrastructure, making the system lightweight and scalable.

Supabase Cloud Services

- **Authentication**: Supabase manages user registration, login, and token-based session handling securely. It supports OAuth2 and JWT for secure identity verification.
- **Cloud Image Storage**: All uploaded images are stored securely in Supabase Buckets, a highly scalable and performant storage solution. This ensures data integrity and allows for future reference or auditing if needed. Supabase serves as the bridge between the backend logic and persistent cloud services, offering both real-time data capabilities and secure file storage in one unified platform.

V. WORKING

The system is designed to provide users with a fast, secure, and intelligent way to determine whether an uploaded image is real or synthetic. By seamlessly integrating cloud-based authentication, frontend interactivity, machine learning, and secure storage, the platform ensures high performance and a smooth user experience. This chapter describes how each component works together, from login to image prediction.

5.1 User Authentication and Image Upload

When users access the system, they are first prompted to either register or log in. This process is securely managed by **Supabase Authentication Services**, which ensure safe and encrypted user access.

Supabase Authentication Features:

- Password Security: Passwords are securely hashed and stored using industry standards.
- JWT Session Management: After login, a JSON Web Token (JWT) is issued to maintain session validity without re-authentication.
- Secure Communication: All user data is transmitted over HTTPS to prevent interception or tampering.

Once authenticated, users are directed to a dashboard where they can interact with the image analysis tools.

Frontend: HTML, CSS, JavaScript

The **frontend** is designed to be lightweight, responsive, and user-friendly. It is developed using:

- **HTML**: To define the structure of the web interface.
- **CSS**: To style components and ensure mobile responsiveness.
- **JavaScript**: To provide interactivity and handle communication with the backend API.

Key Features:

- Drag-and-drop or browse-to-upload image interface
- Real-time form validation and error handling
- Display of predictions, confidence scores, and explanatory feedback

5.2 Image Upload and Cloud Storage

When an image is uploaded, it is first validated on the client side for:

- Format (e.g., JPG, PNG)
- Size and resolution limits
- Corruption or empty files

The validated image is then uploaded to Supabase Buckets, a secure and scalable cloud storage solution.

Supabase Storage Workflow

- **Upload**: JavaScript sends the image to a Supabase storage bucket via signed URL or authenticated API call.
- **Metadata**: Each image is stored with metadata like user ID, timestamp, and filename.
- **Access Control**: Supabase uses fine-grained access rules to ensure only authorized sessions or backend services can access the files.

After upload, a **trigger notifies the backend** (via API or webhook) to begin processing.

5.3 Prediction and result Presentation

Once the LLM API completes its analysis, the result is sent back to the frontend via a structured JSON response.

Frontend Displays:

Prediction: "Real" or "Synthetic"

Confidence Score: A percentage showing how certain the model is (e.g., 87% Synthetic)

Explanation: A user-friendly sentence explaining the result

Visual Element:

- Confidence meters (progress bars)
- Colored tags (green = real, red synthetic) Users can now:

- Upload another image for analysis
- Log out securely, ending

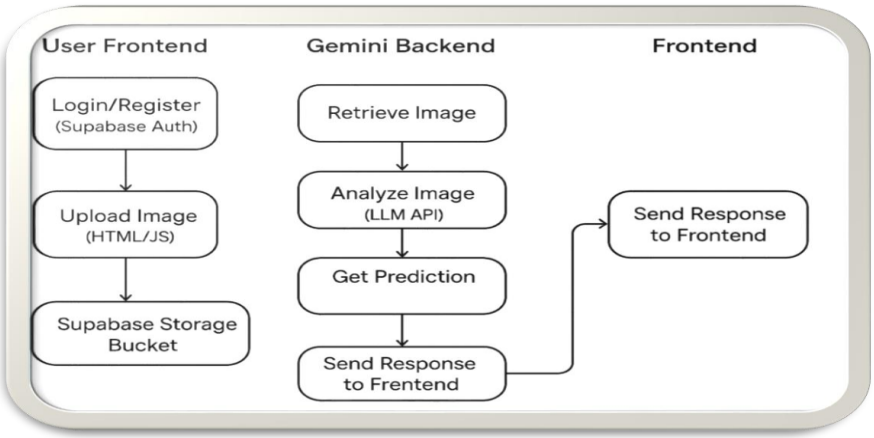

Fig 5.1 System Workflow Diagram

VI. APPLICATION

6.1 Social Media and Content Platforms

Social media platforms like Facebook, Instagram, and Twitter have witnessed a surge in deepfake and manipulated images. The potential consequences of spreading misinformation or harmful content through these platforms can have wide-reaching social, political, and psychological effects. This system can offer several benefits:

- Automatic Flagging of Deepfake Content: The system can be integrated into content moderation tools to automatically flag deepfake images before they are posted or shared. Using AI, it can detect alterations or synthetic images in real time, which helps moderators filter out harmful content at scale.
- Misinformation Prevention: By verifying the authenticity of images shared across these platforms, the system can mitigate the risk of spreading misleading visuals, which are often used to deceive or manipulate public opinion. This ensures the accuracy of visual media circulated online.
- Platform Authenticity and User Trust: The presence of deepfake and AI-generated content undermines user trust in social media platforms. By utilizing this system, platforms can strengthen their commitment to authenticity, reassuring users that the content they encounter is legitimate. Users are more likely to engage with platforms that are proactive in preventing the spread of synthetic images.
- Real-Time Feedback for Users: Platforms can provide users with instant feedback on the authenticity of images, empowering them to make informed decisions when consuming or sharing visual content.

6.2 Journalism and Media Verification

In the realm of journalism, the authenticity of images is essential to maintaining the credibility of news organizations. As the spread of fake news continues to rise, verifying visual content has become a critical task. The proposed system can provide significant support:

- Real-Time Image Verification: Journalists and media organizations can utilize this system to verify the authenticity of images in real-time before they are published. Thiscapability prevents the accidental dissemination of fake or manipulated content, which could severely damage a news outlet's reputation.
- Editorial Decision-Making: With the system in place, editors can make more informed decisions regarding the visual content they publish. Knowing whether an image is real or fake ensures that the editorial process remains unbiased and factual.
- Enhancing Journalistic Integrity: By filtering out synthetically altered or fabricated images, the system helps uphold journalistic integrity. It prevents the manipulation of visual evidence, particularly in sensitive cases, reinforcing public trust in the media.
- Investigative Journalism Support: Investigative journalists who often deal with sensitive and controversial topics can use this system to confirm the authenticity of images before relying on them as evidence in their reporting.

6.3 Legal and Forensic investigation

In legal and forensic contexts, visual evidence plays a pivotal role in investigations. Deepfake images and manipulated photos can be used as fraudulent evidence in court cases, criminal investigations, and cybercrimes. The proposed system offers the following advantages:

- **Validating Court Evidence**: In legal cases, the integrity of visual evidence is crucial. This system can help lawyers, judges, and investigators verify that the images presented in court have not been tampered with, ensuring fair trials and judgments.
- **Assisting Digital Forensics**: Forensic experts tasked with analyzing potential evidence can use the system to quickly identify whether images have been altered or generated synthetically. This is particularly useful in the digital age, where manipulation tools are widely accessible.
- **Combatting Cybercrime and Fraud**: Cybercriminals can use deepfake images to fabricate evidence for fraudulent activities such as identity theft, financial fraud, and defamation. The system can help law enforcement agencies differentiate between real and synthetic images, facilitating the detection and prevention of such crimes.
- **Detecting Evidence in Fraudulent Cases**: The system can be applied to examine images used in fraudulent schemes, particularly those involving altered photos in insurance fraud, impersonation, and other criminal activities.

6.4 education and Awareness Tool

The rise of synthetic image creation tools presents both a challenge and an opportunity for education. This system can play a critical role in raising awareness about the risks and benefits of AI image generation:

- **Educational Demonstrations**: Educational institutions and online platforms can use the system to demonstrate how deepfake and AI-generated images work. This can serve as a powerful tool for teaching students about AI, machine learning, and digital media literacy.
- **Media Literacy Programs**: By helping students and the general public distinguish between real and fake images, the system can promote media literacy, which is essential in an era of rampant misinformation.
- **Awareness Campaigns**: Schools, universities, and non-profits can integrate the system into their educational campaigns to raise awareness about the dangers of manipulated media. This empowers individuals to critically evaluate the authenticity of the content they encounter online.

VII. SNAPSHOT

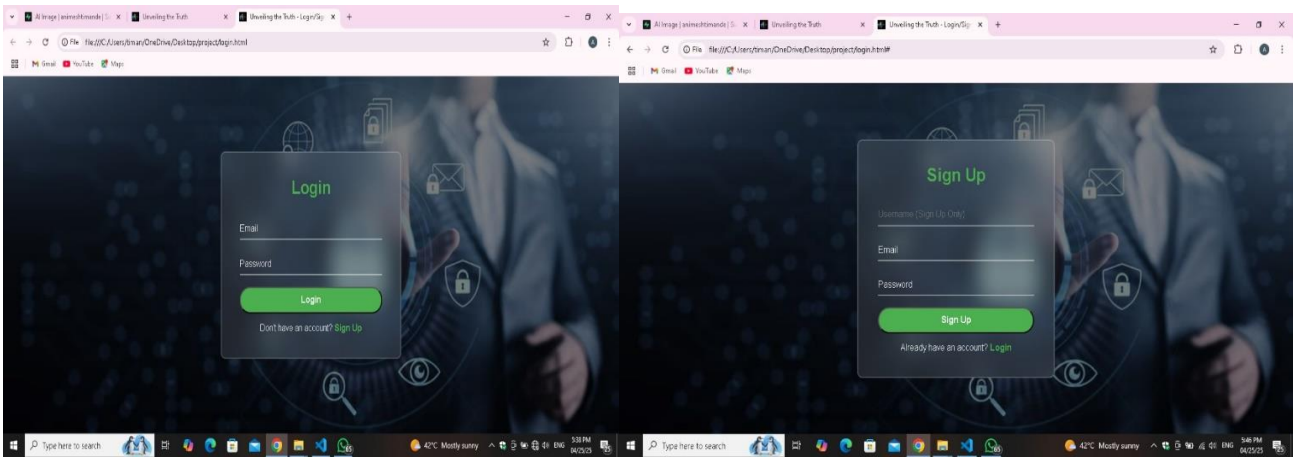

Fig 3: SignUp , login Page

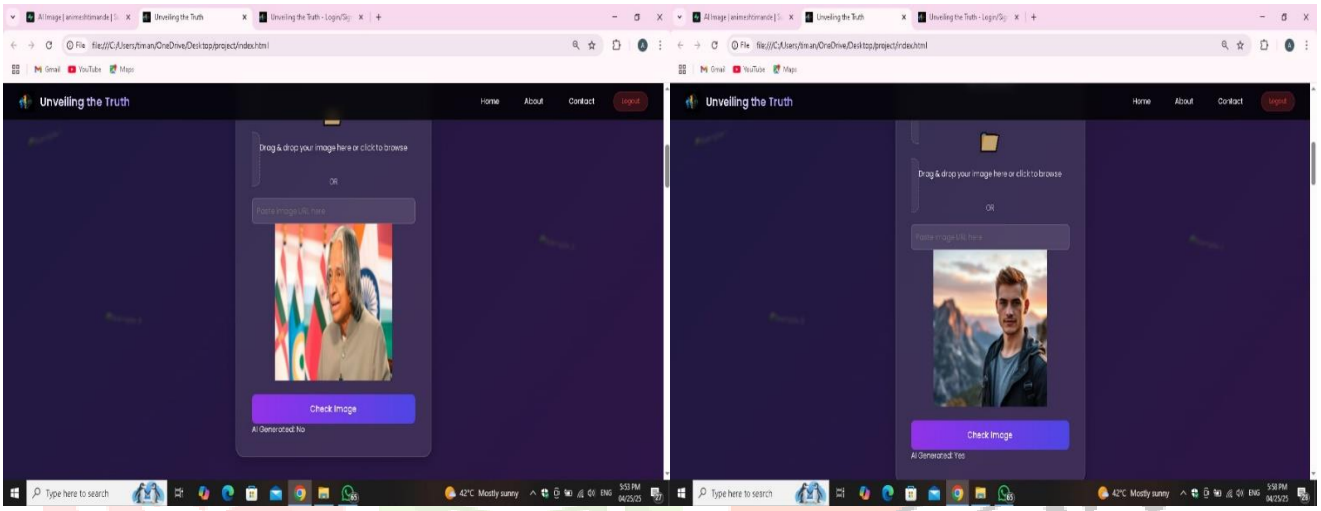

Fig 4: Image Check

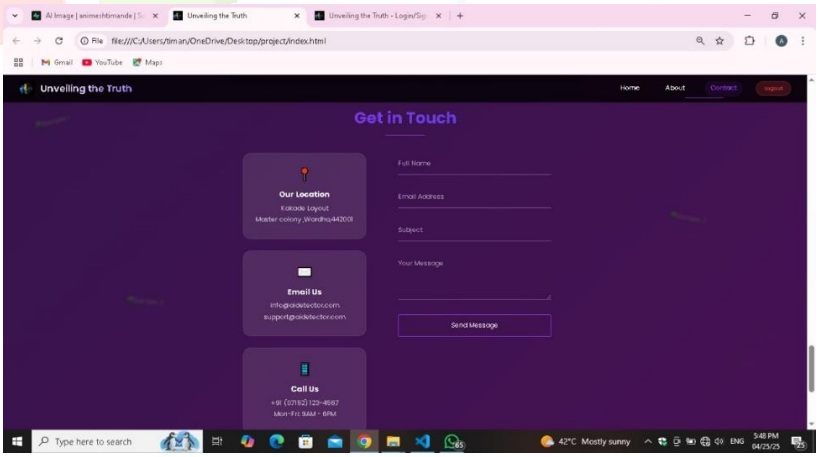

Fig 5: Get In Touch Form

VIII. Future Scope

1. Extension to Video Deepfake Detection
Currently, the project focuses on static image analysis. In the future, it can be extended to detect synthetic content in video formats, which are more commonly used for deepfakes in media, politics, and cybercrime. This involves analyzing temporal inconsistencies, audio-visual mismatches, and motion patterns using models like LSTM, 3D CNNs, or Transformers.

2. Real-Time Detection Capabilities
Integrating the detection model into applications or platforms that can process live images or video streams (e.g., webcams or social media feeds) will enable real-time fake media detection, which is crucial for live broadcasting and social platforms.

3. Multi-Modal Fake Content Detection
The future can also include detection of audio deepfakes and text-based AI-generated content, making the system capable of analyzing multiple modalities of synthetic content. This would make the solution more comprehensive in fighting misinformation.

4. Integration with Browser Extensions or Mobile Apps
Developing lightweight versions of the model that can run on browsers (as extensions) or mobile apps would make it more accessible for everyday users. Such tools could warn users about suspected fake content while they browse the internet or use social media.

IX. Conclusion

In the rapidly evolving digital landscape, the ability to distinguish real from synthetic images has become a critical concern, as the proliferation of advanced AI technologies, such as deepfakes and generative models, continues to blur the lines between authentic and manipulated content. This project aimed to explore and develop AI-driven solutions for image verification, utilizing cutting-edge techniques in deep learning, image forensics, and blockchain technology to ensure the integrity of digital images. Our findings demonstrate that while significant progress has been made in developing robust verification systems, challenges remain, especially as generative technologies continue to improve.

This project, "Unveiling the Truth: Exploring AI Solutions to Identify Real vs Synthetic Images," demonstrates the potential of Large Language Models (LLMs), particularly leveraging Gemini technology, in the domain of deepfake and synthetic image detection. By harnessing the advanced reasoning and multimodal capabilities of LLMs, our system analyzes visual data with contextual understanding, surpassing traditional image classification techniques. The integration of this intelligent backend with a user-friendly web interface and cloud services like Supabase ensures both accessibility and scalability. This innovative approach not only enhances the accuracy of identifying manipulated content but also sets a new direction for future research in AI-powered digital authenticity verification.

X. Reference

[1]. Zhang, X., & Dong, X. (2020). Deep learning for detecting image forgery: A survey. Pattern Recognition, 105, 107294. https://doi.org/10.1016/j.patcog.2020.107294

[2]. Dolhansky, B., et al. (2020). The deepfake detection challenge. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1-9. https://doi.org/10.1109/CVPR42600.2020.00010

[3]. Yli-Huumo, J., Ko, D., Choi, S., & Park, S. (2016). A survey of blockchain: A comprehensive review of applications, challenges, and opportunities. Proceedings of the 2016 IEEE 13th International Conference on Embedded Software and Systems (ICESS), 18-24. https://doi.org/10.1109/ICESS.2016.17

[4]. Ali, M. M., & Vasilenko, D. (2018). Blockchain technology and its applications in image provenance. International Journal of Computer Science and Network Security, 18(10), 37-45. https://doi.org/10.22937/IJCSNS.2018.18.10.37

[5]. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. Proceedings of the International Conference on Machine Learning (ICML), 1-9. https://arxiv.org/abs/1412.6572

[6]. Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. Proceedings of the IEEE Symposium on Security and Privacy (SP), 39-57. https://doi.org/10.1109/SP.2017.49