

Project Proposal

Himang Chandra Garg Dasari Sai Harsh Nishil Agarwal Piyush Narula
Indraprastha Institute of Information Technology, Delhi
{himang22214, dasari22144, nishil22334, piyush}@iiitd.ac.in

1. Motivation

Our project seeks to address the critical issue of detecting deepfake images and AI-generated content, which have become increasingly sophisticated and challenging to distinguish from real media. The goal is to enhance the ability to discern genuine content from fabricated images, thereby reducing the potential for misinformation.

1.1. Why this project?

The recent rise in AI content, particularly deepfake images and videos, has raised serious concerns about the integrity of information online. This synthetic media is often indistinguishable from real content, hence making it very difficult for an average person to tell the truth—an aspect with large implications in public trust, privacy, and security. Enhancements in detections are required to counter the spread of misinformation, secure privacy, and maintain the integrity of digital media toward a more secure and trustworthy online environment.

1.2. How did you think about this?

We visualized this project after many problems of AI-generated content in work and personal experience. In search of an answer for that gap, we observed that no website or app formally available is solving the challenge effectively. This realization motivated us to create our own system—one that could effectively detect and reduce the risks associated with AI-generated content (specifically images).

2. Related work

2.1. Detecting Fake Images Using Machine Learning

This article utilizes methods such as Convolutional Neural Networks (CNNs) for pattern recognition, feature extraction techniques for identifying statistical properties and textures, and conventional image forensics methods like digital watermarking and error level analysis.

2.2. Work 2

2.3. Work 2

3. Timeline

4. Individual Tasks

4.1. Himang Chandra Garg

4.2. Dasari Sai Harsh

4.3. Nishil Agarwal

4.4. Piyush Narula

5. Final Outcome

5.1. What are you expecting from this project?

Our first main goal is to establish a high level of accuracy in the model when it can distinguish between real and AI-generated pictures even in the case of deep fakes. We hope to use deep learning and machine learning techniques to find patterns and irregularities in this synthetic content that partition it from real content. This model, therefore, should have very high precision and recall and hence should be dependable. Additionally, we plan to implement a basic front-end interface that will make this detection system accessible, addressing the current lack of formal tools or platforms dedicated to this purpose.

5.2. What do you want to contribute to this idea?