A Synopsis

of

Minor Project [CC3270]

**Real-Time Detection of AI-Generated   
Images and Deepfake**

Submitted to Manipal University Jaipur

towards the partial fulfillment of the requirement for the award of the degree of

**BACHELOR OF TECHNOLOGY**

in

Computer and Communication Engineering

Jan – May 2025

By:

Vansh Giri

229303045



Under the guidance of:

Dr. Vaishali Yadav and Dr. Arpit Kumar Sharma

**Department of Computer and Communication Engineering**

**Manipal University Jaipur**

**INTRODUCTION**

The rapid advancements in generative AI have led to significant breakthroughs in image synthesis, enabling the creation of highly realistic deepfake images. While these technologies have beneficial applications in entertainment, media, and creative fields, they also pose serious ethical and security risks. AI-generated images can be misused for misinformation, identity fraud, and digital manipulation, making it crucial to develop reliable detection mechanisms.

This project focuses on real-time detection of AI-generated images and deepfakes. With the rise of sophisticated deep learning models such as GANs (Generative Adversarial Networks), distinguishing between real and synthetic images has become increasingly challenging. The ability to detect such manipulations is essential for ensuring digital content authenticity, preventing misinformation, and maintaining trust in online media.

To address this issue, I have chosen a dataset from Kaggle that includes both real and AI-generated images. The dataset consists of various deepfake samples generated using advanced AI techniques, allowing for effective model training and evaluation. By leveraging deep learning models such as Convolutional Neural Networks (CNNs) and Transformer-based architectures, this project aims to develop a system capable of identifying AI-generated images with high accuracy and efficiency.

A key aspect of this project is its real-time detection capability, which is critical for applications such as media forensics, social media moderation, and cybersecurity. The project will culminate in the development of a web-based tool that can analyse image inputs and determine their authenticity. This user-friendly interface will make AI-generated image detection accessible to a broader audience, including journalists, organizations, and law enforcement agencies.

By choosing this as my minor project, I aim to contribute to the growing field of AI security and digital forensics. The increasing prevalence of deepfake technology makes it imperative to develop countermeasures that can help detect and mitigate potential threats. This project will not only enhance my understanding of machine learning and deepfake detection but also address a real-world challenge with significant societal implications.

**MOTIVATION**

The widespread availability of AI-generated image technology has introduced both opportunities and threats to digital media. While synthetic images can enhance creative expression and digital content generation, they also pose risks related to misinformation, identity fraud, and privacy breaches. The misuse of deepfake technology in media manipulation and deceptive campaigns underscores the urgent need for robust detection mechanisms.

My motivation for this project stems from a strong interest in AI ethics and cybersecurity. As deepfake image technology becomes more advanced, it is essential to develop real-time solutions that can detect and counteract fraudulent visual content. This project provides an opportunity to work with state-of-the-art deep learning models, gain hands-on experience in computer vision techniques, and contribute to a pressing global issue.

Furthermore, the challenge of building a real-time AI-generated image detection system aligns with my goal of applying theoretical knowledge to practical, impactful solutions. The development of a web-based detection tool ensures that this research has real-world applications, making it accessible to individuals, media organizations, and cybersecurity teams that need to verify the authenticity of digital images.

By working on this project, I aim to bridge the gap between AI advancements and security concerns, ensuring that technological progress does not come at the cost of ethical risks. This project is not only a technical endeavour but also a step towards making AI safer and more responsible.

**PROBLEM STATEMENT**

With the rise of AI-driven image synthesis and deepfake generation, there is an increasing risk of malicious actors using these technologies for fraudulent activities such as identity theft, misinformation, and media manipulation. Current detection mechanisms are either ineffective in real-time scenarios or require significant computational resources, making them impractical for widespread deployment.

The primary challenge is to develop a real-time AI-generated image detection system that can accurately distinguish between real and AI-generated images. The solution should be scalable, efficient, and accessible, ensuring that individuals, media organizations, and cybersecurity professionals can verify the authenticity of digital images without requiring specialized expertise.

**Objectives**

1. Develop a machine learning-based model capable of detecting AI-generated images and deepfakes in real-time.
2. Utilize a dataset comprising real and fake images to train and evaluate the model.
3. Implement a web-based tool that allows users to analyse images for deepfake detection.
4. Ensure the model is lightweight and optimized for real-time performance.
5. Assess the model’s accuracy, efficiency, and robustness against evolving deepfake techniques.

**Pros & Cons of Existing Methods**

**Pros**

* Traditional forensic analysis methods, such as error level analysis (ELA) and metadata inspection, provide some effectiveness in detecting AI-generated images.
* Deep learning approaches, such as Convolutional Neural Networks (CNNs) and Vision Transformers, have shown promising results in identifying synthetic images.
* Existing forensic tools can analyse artifacts, inconsistencies, and texture anomalies to flag potential deepfakes.

**Cons**

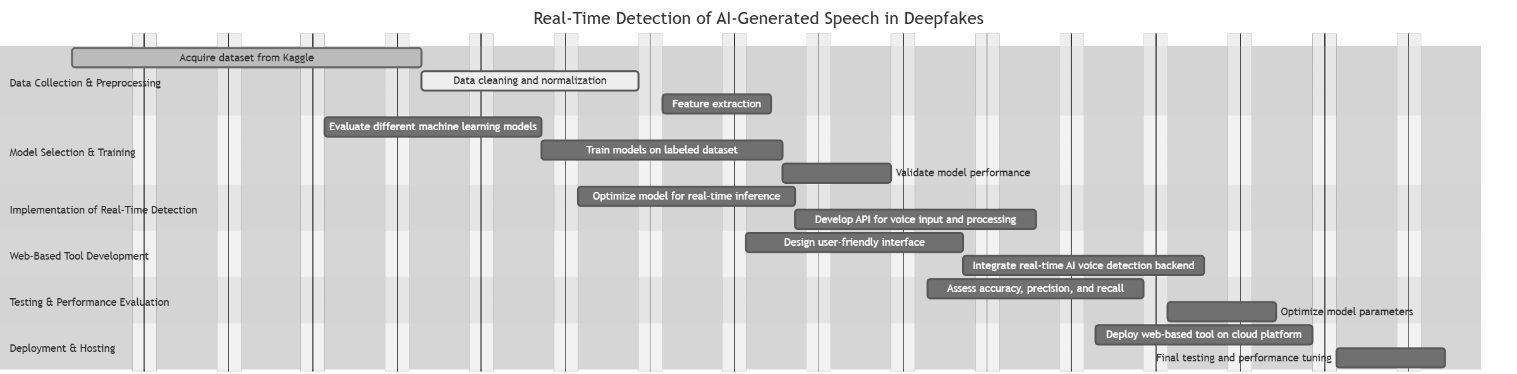
* Many current detection systems are computationally expensive and require powerful hardware.
* Some models rely heavily on supervised learning, making them vulnerable to adversarial attacks and poor generalization.
* The evolving nature of AI-generated images means that detection methods must continuously adapt to new synthesis techniques.
* A lack of publicly available real-time detection tools makes it difficult for non-experts to verify image authenticity.

By addressing these limitations, this project aims to create a practical and effective solution for real-time AI-generated image detection, contributing to the broader effort of safeguarding digital content from deepfake threats.

**METHODOLOGY**

To achieve the objectives of this project, the following steps will be followed:

1. Data Collection & Preprocessing
   * Acquire a dataset from Kaggle, containing real and AI-generated images.
   * Perform data cleaning, resizing, augmentation, and normalization.
2. Model Selection & Training
   * Evaluate various deep learning models (CNNs, Vision Transformers, EfficientNet, ResNet).
   * Train models using labelled datasets and validate performance with test data.
3. Implementation of Real-Time Detection
   * Optimize the model for real-time inference with minimal latency.
   * Develop an API for image input and processing.
4. Web-Based Tool Development
   * Design and implement a user-friendly web interface for image analysis.
   * Integrate real-time AI-generated image detection with backend processing.
5. Testing & Performance Evaluation
   * Assess the accuracy, precision, recall, and F1-score of the detection system.
   * Optimize model parameters for improved real-time performance.
6. Deployment & Hosting
   * Deploy the web-based detection tool on a cloud platform.
   * Conduct final testing and performance tuning before release.



Gantt Chart of Methodology

This structured methodology ensures a systematic approach towards achieving the project's objectives and delivering a functional AI image detection system.

**FACILITIES REQUIRED**

**Software Requirements:**

* Python (with libraries such as TensorFlow, PyTorch, OpenCV, and Scikit-learn)
* Flask/Django for web backend development
* React.js/HTML/CSS for frontend development
* MongoDB/MySQL for database management (if required)
* Cloud services (AWS, Google Cloud, or Azure) for hosting
* Jupyter Notebook for development and experimentation

**Hardware Requirements:**

* No specific hardware requirements, as the project will be conducted on my personal computer.
* If additional computational resources are needed, access to lab equipment or cloud-based GPUs will be considered.

**BIBLIOGRAPHY & REFERENCES**

Dataset: [CIFAKE: Real and AI-Generated Synthetic Images](https://www.kaggle.com/datasets/birdy654/cifake-real-and-ai-generated-synthetic-images)

Research Papers:

* [Detecting Deepfake Images Using Deep Learning Techniques and Explainable AI Methods](https://cdn.techscience.cn/ueditor/files/iasc/TSP_IASC-35-2/TSP_IASC_29653/TSP_IASC_29653.pdf)
* [Deepfake attribution: On the source identification of artificially generated images - Khoo - 2022 - WIREs Data Mining and Knowledge Discovery - Wiley Online Library](https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1438)