



AI-Generated Media Detection System

Supervised by Dr. **Eman Abdel-Latef**

Program Name	
<input type="checkbox"/> Computer Science <input type="checkbox"/> Information Systems	<input type="checkbox"/> Scientific Computing <input checked="" type="checkbox"/> Artificial Intelligence
English Title	AI-Generated Media Detection System
Arabic Title	نظام الكشف عن الوسائط المولدة بالذكاء الاصطناعي
Project Team	
1. Romani Nasrat Shawqi Gerges	1. روماني نصرت شوقي جرجس
2. Ahmed Mohamed Ali Abo el-Kassem	2. احمد محمد علي ابو القاسم
3. Zeyad Elsayed Abdel-Azim Ali	3. زياد السيد عبدالعظيم علي يوسف
4. Sara Reda Moatamed Hassan	4. ساره رضا معتمد حسن
5. Reham Moustafa Ali Abdel-Moati	5. ريهام مصطفى علي عبد المعطى
6. Rawan Abdel-Aziz Ahmed Mahmoud	6. روان عبدالعزيز أحمد محمود
7. Abd-Allah Mohamed Abdel-monem	7. عبدالله محمد عبدالمنعم عبدالفتاح
8. Mohannad Ayman salah Abdel-fattah	8. مهند ايمن صلاح عبدالفتاح
9. Mohamed Abd-Allah Abdel-salam Abdel-Dayem	9. محمد عبدالله عبدالسلام عبدالدايم
Project Supervisor	
Faculty member	Dr. Eman Abdel-Latef
Assistant	
Teaching Assistant	Eng. Mohamed Mostafa

Contents

1.	Introduction.....	4
2	Project Overview:	5
2.1	Problem definition:.....	5
2.2	Objective.....	5
3.	Key Features:.....	6
4.	Scope of Work	7
5.	Datasets.....	8
5.1	images Data	8
5.2	Text Data.....	8
5.3	Audio Data.....	8

1. Introduction

Artificial Intelligence (AI) has evolved into a powerful tool for generating voice-based conversations, articles, and facial images. However, with this power comes the potential for misuse, leading to harm and negative consequences. The aim of this project is to develop a sophisticated system capable of detecting AI-generated content in the forms of audio, images, and text. By implementing modern technologies, we seek to address the challenges associated with unregulated AI use and contribute to the establishment of ethical boundaries.

2. Project Overview:

2.1 Problem definition:

The rapid advancements in **Generative adversarial network (GAN)** technology in generating images, voice-based conversations, articles, and facial images. have raised concerns regarding its potential misuse, both from illegal activities causing harm and unintended negative consequences within educational contexts

- AI has the potential to be misused in illicit ways, harming individuals. Such as the blackmail operations with pictures or videos that were Generated by with AI
- Students are using AI for scientific tasks, inadvertently impacting education negatively.
- Some AI agents may acts like Humans by generating different formats of media to access websites or services, this causes some problem.

This necessitates establishing boundaries for using generative AI.

2.2 Objective

Our primary objective is to Develop AI-based System helps Humanity from damages by modern GAN systems. our System can:

- **Discerning** between contents (image/ text/audio) **generated** by **AI** and content of **human** origin.
- Contribute to the ethical use of AI technology by mitigating potential harm.
- Safeguard users from the deleterious effects of unchecked AI capabilities.

3. Key Features:

key features of our web application system for detecting AI-generated:

1. Detection of Ai generated Contents:

this is the core feature of our system, it gives use ability to check if content is real or generated by AI , here we work with 3 data formats : Images, Text, and Audios

2. Registration Functionalities:

Users can login or sign up. This give user more usability of system features.

3. Contribution of the user to collect data:

this option allows the user to provide samples of data with labels to help us collecting more data.

4. Feedback Mechanism:

users can report false positives or false negatives,
We use this feedback to continuously refine and enhance.

4. Scope of Work

- **Text Detection:** Implement Natural Language Processing (NLP) techniques to identify patterns and linguistic cues indicative of AI-generated text. Explore pre-trained language models like GPT-3 and BERT for effective detection.
- **Image Analysis:** Utilize computer vision algorithms to analyze image content and identify anomalies typical of AI-generated images. This may involve examining pixel patterns, metadata, and other relevant features.
- **Audio Analysis:** Apply signal processing and machine learning techniques to analyze audio files, detecting anomalies that may indicate AI-generated voices or manipulated audio content.
- **Cross-Format Verification:** Implement a cohesive system that can cross-verify results from different formats to increase the overall accuracy of AI-generated media detection.
- **User Interface:** Develop an intuitive user interface that allows users to submit media for analysis and presents clear results, indicating the likelihood of AI generation.
- **Continuous Learning:** Incorporate a machine learning pipeline that continuously learns and adapts to new AI-generated techniques, ensuring the system remains up-to-date and effective over time.

5. Datasets

5.1 images Data

✧ AI-Art Bench

AI-ArtBench is a dataset that contains **180,000+** art images. **60,000** of them are human-drawn art that was directly taken from ArtBench-10 dataset and the rest is generated equally using Latent Diffusion and Standard Diffusion models. The human-drawn art is in **256x256** resolution and images generated using Latent Diffusion and Standard Diffusion have 256x256 and 768x768 resolutions respectively.

Real vs AI generated human faces

Fake Human(6873 images) , Real Human(7000 images).

We have some videos datasets we may use it as we can use it to get images by splitting videos into single frames

- **celeb-df**
- **ff++ dfdc (deepfake detection challenge)**
- **fakeAVceleb**

5.2 Text Data

✧ -DAIGT | External Dataset

dataset for real and ai generated text

5.3 Audio Data

✧ DEEP-VOICE: DeepFake Voice Recognition:

This dataset contains examples of real human speech, and DeepFake versions of those speeches by using Retrieval-based Voice Conversion.