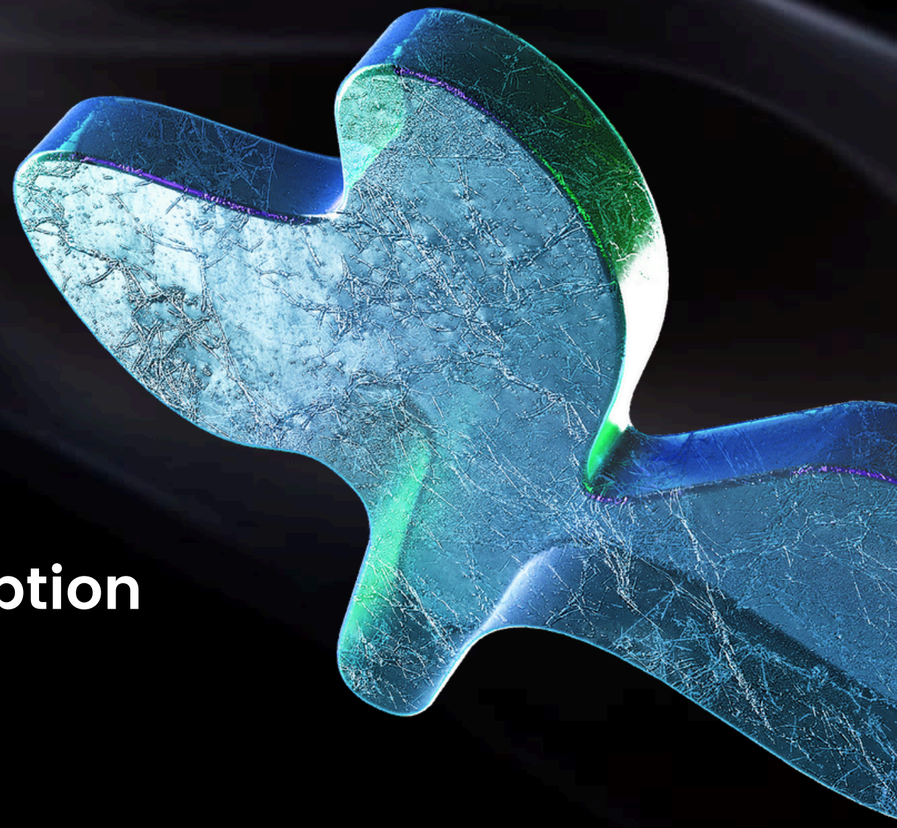
PROJECT - II

PixelTruth

Empowering Authenticity in a World of Digital Deception

GROUP 05 - AI VS REAL IMAGES



Our Team

Radhika Bhati

Deepanshu Aggarwal

Rohit

Harshita Rupani

Table of Content

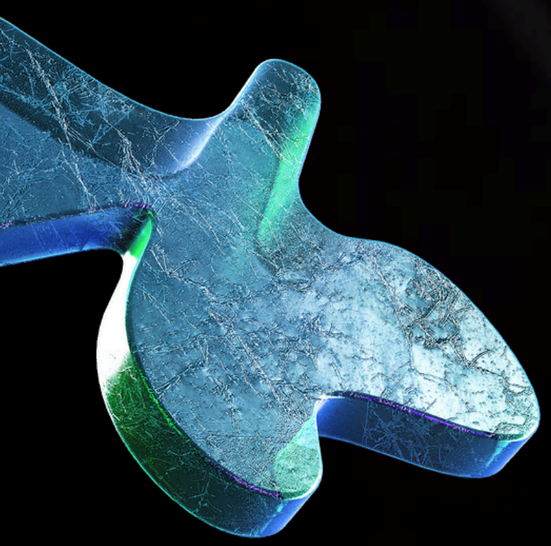
What all we will present?

Problem Statement

In an era where artificial intelligence can produce images that blur the lines between reality and fiction, differentiating between authentic photographs and AI-generated ones has become increasingly difficult. This poses significant challenges. AI-generated images can be manipulated to fabricate events and spread false narratives, leading to serious ethical implications. Trust in visual content is essential, yet the prevalence of AI-generated images undermines this trust, making verification challenging for users. Addressing these issues requires a reliable technology solution to differentiate between real and AI-generated images, promoting ethical standards and truthfulness in digital media.

Insights from Expert

Identifying Key Issues



How do you know whats real and what isnt?
and Does it even matter anymore?

Does it even matter that photos are AI or not AI?

TITLE - AI Images Vs Real Photos
click here to see full video

Do we need something that not only restricts
but penalises misuse of AI-generated images?

Role of the media is crucial, in clearly labelling when something is manipulated or fake. As well as role of technology which can give us digitial watermarks showing whats real or fake.

TITLE - How to differentiate between AI-generated images and videos from real ones
click here to see full video

There is currently no technology available to determine if a video is machine-generated or deepfake

The timing of deepfake detection and takedown is critical due to the rapid spread of manipulated videos.

Deepfake Dangers and Debate:
The Role of AI in Politics & Media
click here to see full video

Objective

The primary objective of our project, "PixelTruth" is to develop a robust technology solution capable of accurately distinguishing between authentic photographs and AI-generated images. This solution aims to address the critical issues presented by the proliferation of AI-generated images as discussed in previous slides:

MISINFORMATION AND DECEPTION

By enabling the differentiation between real and AI-generated images, we aim to reduce the spread of misinformation and prevent the fabrication of events and false narratives.

ETHICAL CONCERNS

Our objective is to uphold ethical standards in digital media by providing a tool that helps identify and reduce the misuse of AI-generated images, thereby reducing the potential for harm, such as the creation of fake identities and forged evidence.

TRUST AND VERIFICATION

We seek to restore trust in visual content by offering users a reliable means of verifying the authenticity of images encountered online. By promoting transparency and truthfulness, our solution aims to enhance trustworthiness in the digital landscape.

Flow Of Data Pre-processing

- Acquired CIFAKE dataset consisting of 60,000 AI-generated and 60,000 real images from CIFAR-10.
- Dataset categorized into REAL (CIFAR-10) and FAKE (AI-generated) classes.
- Supplemented dataset with 660 images via web scraping, encompassing both AI-generateD (330) and real photographs(330).
- Due to hardware limitations, reduced training dataset by 70% and testing dataset by 90%.
- Revised dataset: 30,000 images for training (15,000 real, 15,000 AI-generated) and 2,000 images for testing (1,000 real, 1,000 AI-generated).
- Changed the names of images, replacing space with “_”

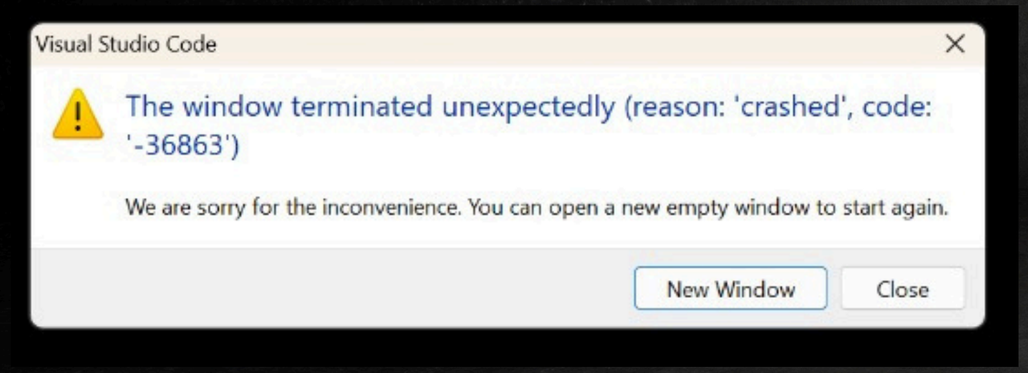
Flow Of Data Pre-processing

- Conducted preprocessing: resized all images to consistent size and normalized pixel values.
- Resizing ensures uniform dimensions for efficient processing and analysis.
- Pixel normalization standardizes pixel values for optimized model performance.
- We succesfully trained the model and hence decided to increase the dataset size.
- Tried to re-train the model on 70,000 images (35,000 real, 35,000 AI-generated)
- Faced Hardware limitations yet again and decided to reduce the size.
- Successfully re-trained with almost 50,000 images (25,000 real, 25,000 AI-generated) as well as 660 web-scraped images.
- Also, compiled a set of real images of various indian celebrities from google and generated a set of AI-generated images of the same celebrities using leonardo.ai , added those in the dataset as well.

Flow Of Data Pre-processing

- SOME OF THE ERRORS WE FACED DUE TO HARDWARE LIMITATIONS:

Python ran out of memory (RAM) or encountered other resource limitations. This happens if we work with large datasets or complex models.



Training a deep learning model can be computationally intensive and may require a significant amount of memory (RAM) and processing power (CPU/GPU). If the system runs out of resources during training, it can lead to a crash.

Literature Review

PAPER TITLE	AUTHOR	KEY FINDINGS
Quantifying the Performance Gap between Real and AI-Generated Images (2023)	Shivani Atul Bhinge, Piyush Nagpal	By analyzing various performance metrics like accuracy, precision, recall, and F1 score, this research paper aim to identify disparities between the two image types. Insights into the strengths and limitations of AI-generated synthetic images are provided, along with guidance for incorporating them into real-world applications. The study contributes to advancing computer vision by enhancing understanding of the suitability and reliability of AI-generated synthetic images, using the CIFAKE dataset for model training .
GenImage: A Million-Scale Benchmark for Detecting AI-Generated Images (2024)	Mingjian Zhu et al.	This research paper introduces the GenImage dataset, designed for training and evaluating detectors to distinguish between AI-generated fake images and real ones. It conducts a thorough analysis of dataset characteristics and proposes evaluation tasks resembling real-world conditions, such as cross-generator image classification and degraded image classification. By assessing detector performance across different image synthesis techniques and on degraded images, the study aims to advance detector development tailored for AI-generated image detection.

Literature Review

PAPER TITLE	AUTHOR	KEY FINDINGS
Harnessing Machine Learning for Discerning AI-Generated Synthetic Images (2023)	Yuyang Wang, Yizhi Hao, Amando Xu Cong	The research addresses the challenge of identifying AI-generated synthetic images in digital media using machine learning techniques. By enhancing model accuracy through transfer learning and employing the CIFAKE dataset, which contains both real and fake images, it aims to improve synthetic image detection precision. Comparative analysis with traditional methods, like Support Vector Machine (SVM), evaluates the effectiveness of the optimized models.
CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images (2023)	JORDAN J. BIRD, AHMAD LOTFI	This research paper aims to distinguish between real and AI-generated photographs by: (a) Introducing the CIFAKE dataset, comprising 120,000 images, for research purposes; (b) Developing a computer vision method to enhance recognition of AI-generated images; (c) Advocating for Explainable AI to understand image recognition processes and visualize significant features. It addresses societal and ethical concerns regarding image authenticity amid rapid advancements in AI-generated image generation.

Methodology

DATA PREPARATION:

- Loaded the training and testing images from the specified directories.
- Resized the images to a fixed size (48x48 pixels) and normalized the pixel values to the range [0,1].
- Shuffled the training and testing data to avoid any bias in the model.

MODEL ARCHITECTURE:

- Constructed a Convolutional Neural Network (CNN) model using TensorFlow and Keras.
- The model consists of convolutional layers followed by max-pooling layers and dropout layers to prevent overfitting.
- The final layers include dense layers with ReLU activation functions, and a sigmoid output layer for binary classification.

MODEL TRAINING:

- Compiled the model with the Adam optimizer and binary cross-entropy loss function.
- Trained the model using the training data for 15 epochs, validating the performance on the testing data after each epoch.

MODEL EVALUATION:

- Saved the trained model for future use and load it to evaluate its performance on the testing data.
- Compute evaluation metrics such as accuracy, precision, recall, and F1-score using the classification report from scikit-learn.

FINAL RESULT:

- Plot the training and testing loss and accuracy curves to visualize the model's performance over epochs.
- Model Deployed on a website to effectively predict the images using Streamlit.

Results

AI Image Type Detector

Upload an image file...

Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG
Browse files

22lalu-yadav.jpg 25.2KB

Enter the URL of the image:

Detect Image Type

The given image is: Real

AI Image Type Detector

Upload an image file...

Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG
Browse files

Enter the URL of the image:

https://resize.indiatvnews.com/en/resize/newbucket/1080_-/2023/04/untitled-design-2023-04-09t1228

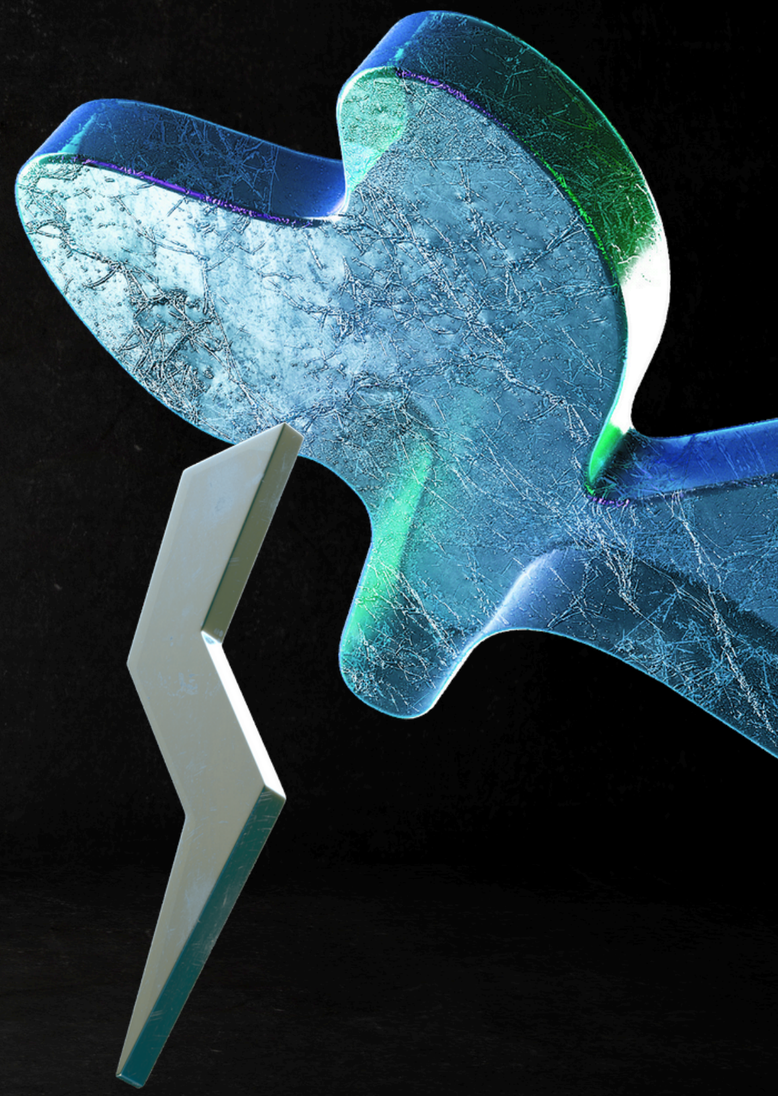
Detect Image Type

The given image is: AI Generated

Future Outlook

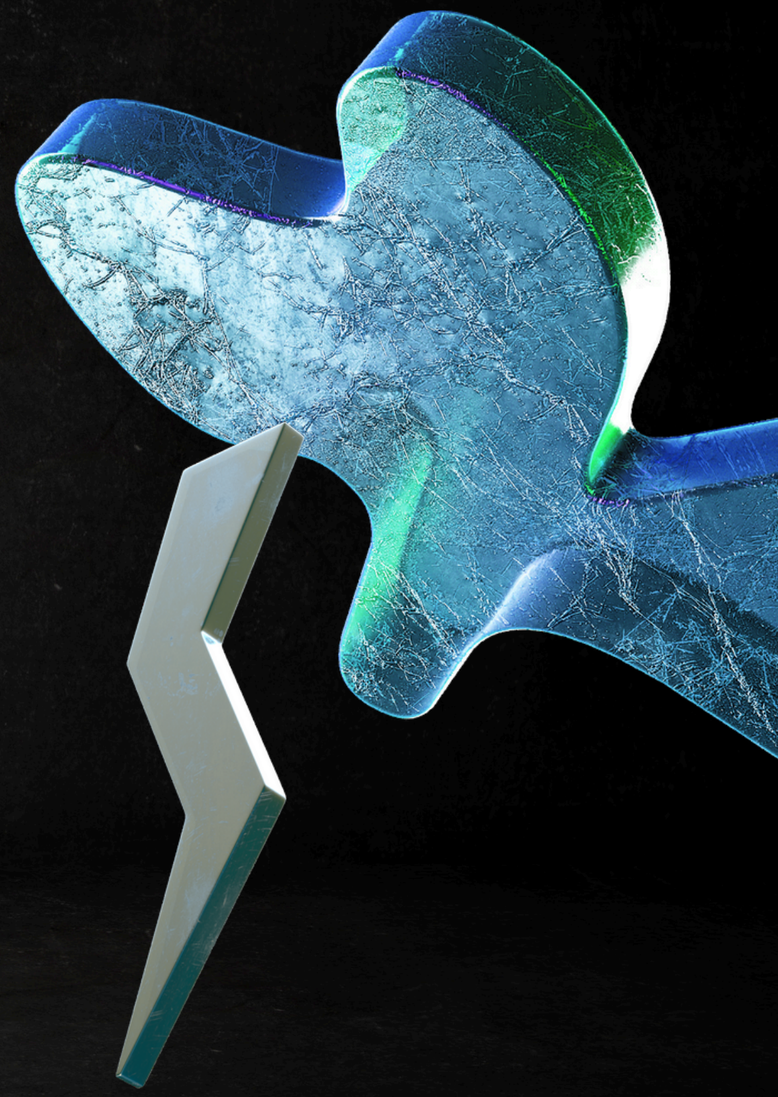
Our next steps

Broaden the model's scope to encompass not only images but also other form of media such as videos, audio recordings, and text, enabling comprehensive detection of AI-generated content across diverse modalities. This holistic approach would ensure a more robust solution for identifying synthetic media, promoting greater trust and reliability in digital content authentication efforts.



Conclusion

In summary, "PixelTruth" emerges as a pivotal step towards addressing the challenges posed by AI-generated images, fostering trust, and upholding ethical standards in digital media. By crafting a comprehensive methodology and addressing key objectives, our project signifies a commitment to mitigate the rise of misinformation and safeguard against ethical breaches in digital media. Through continuous innovation and a forward-looking approach, we aspire to increase trust and reliability in content authentication, laying a foundation for a more transparent and accountable digital ecosystem.



Thank You

GROUP 05 - CSE 04