# DETEK-AI: A Web-based Deepfake Detection System

Dr. Shwetambari Borade
*Department of Cyber Security Shah & Anchor Kutchhi Engineering College Mumbai, India*
shwetambari.borade@sakec.ac.in

Aabha Wagh
*Department of Cyber Security Shah & Anchor Kutchhi Engineering College Mumbai, India*
aabha.wagh15551@sakec.ac.in

Vaidehi Salvi
*Department of Cyber Security Shah & Anchor Kutchhi Engineering College Mumbai, India*
vaidehi.salvi16023@sakec.ac.in

Drashti Nagda
*Department of Cyber Security Shah & Anchor Kutchhi Engineering College Mumbai, India*
drashti.nagda15764@sakec.ac.in

Parth Savla
*Department of Cyber Security Shah & Anchor Kutchhi Engineering College Mumbai, India*
parth.savla16074@sakec.ac.in

***Abstract*– AI-driven deepfake bots pose a serious risk to society, security, and trust, which necessitates the development of reliable techniques to detect and mitigate their threat, as their proliferation leads to disinformation, deceit, and harm. Deepfake detection systems can improve cybersecurity, safeguard digital identities, mitigate misinformation, protect reputations of individuals and organizations, assure ethical use of AI technology, empower people and organizations, support regulatory efforts, and contribute to a safer digital ecosystem. Detek AI is a frontrunner in innovative technology aimed at combating deepfake content, safeguarding individuals and organizations from misinformation, and upholding the integrity of online platforms.**

***Keyword— Deepfake Technology, Deepfake Detection, Convolutional Neural Networks, Artificial Intelligence bots, Machine Learning Artificial Intelligence, Face Morphing Detection, Deep Learning, Benford's Law.***

## I. INTRODUCTION

In recent years, the rapid development of AI and machine learning technology has led to the emergence of deepfake AI bots, posing a significant risk to society, security, and trust in the digital sphere. The prevalence of deepfakes has resulted in widespread deceit, disinformation, and potential harm, necessitating the urgent development of reliable techniques to detect and mitigate this threat. Deepfakes are AI-generated media that replace an individual's face or manipulate their speech in a deceptive manner [1].

Detecting deepfakes is crucial for safeguarding cybersecurity, preserving digital identities, combating misinformation, protecting the reputations of individuals and organizations, promoting ethical AI use, empowering individuals and organizations, supporting regulatory efforts, and contributing to a safer digital ecosystem [2]. Deepfake detection systems play a pivotal role in achieving these goals.

Detek AI, a web-based deepfake detection tool, is at the forefront of innovative approaches to combat the rising threat of deepfake content in the digital sphere. The groundbreaking detection system not only safeguards individuals from misinformation but also upholds the integrity of online platforms, establishing trust in an era where discerning truth from falsehood is paramount.

## II. LITERATURE REVIEW

The current era has seen a rapid increase in Deepfake technology, making it more challenging to discern between actual material and faked content. As technology advances, it has the ability to upend many facets of society, including politics, the media, and interpersonal relationships.

According to the assessment, deepfake creation techniques have evolved over time, combining speech synthesis, full-body movements, and facial changes. The emergence of these multimodal deepfakes highlights the need for comprehensive detection techniques by making detection more challenging.

The research articles emphasize the significance of deepfake detection and provide a variety of strategies for resolving the problem. Some articles offer novel approaches including cyber vaccination, deep-CNN architectures, and blockchain technology, while others concentrate on text-based detection utilizing deep learning models and word embeddings. Every method shows promise, achieving high accuracy scores in detecting deepfakes in text, pictures, and videos, among other formats. In order to achieve accurate deepfake detection, the papers stress the need of feature extraction, model selection, and dataset construction. It is imperative to take into account the constraints and modifications associated with each methodology, as their efficacy may vary contingent upon the particular deepfake type and dataset employed.

All things considered, these publications provide insightful information and significant progress against the proliferation of deepfakes by providing a variety of useful methods for recognizing and detecting faked content.

## III. PROPOSED SYSTEM

With the help of this survey, the paper further proposes the DETEK-AI system. It is an application with the following features:

- **Input:** Specify the path to the folder containing the dataset of images (image_folder).
- **Preparation**: Import necessary libraries (e.g., os, cv2, numpy, matplotlib, scipy). Define utility functions (get_first_digit, benfords_law_test, analyze_image, main).
- **Load Image Dataset and Analyze:** Iterate through each image file in the specified folder.
- **For each image:** Load the image using OpenCV (cv2.imread). Flatten the grayscale image to extract features. Analyze the image using Benford's Law (benfords_law_test function).
- **Deepfake Detection Threshold:** Set a threshold value (threshold) to determine whether an image is classified as a potential deepfake based on the p-value.
- **Display Results:** Iterate through the analyzed images and print the results. If the p-value is below the threshold, mark the image as a potential deepfake; otherwise, label it as likely genuine.
- **Handle Exceptions:** Implement error-handling mechanisms for issues such as image loading failures or division by zero errors.

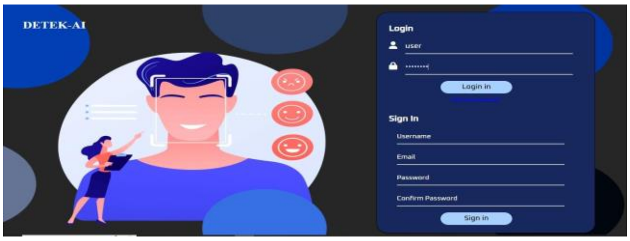


Fig3.1 login page of DETEK-AI

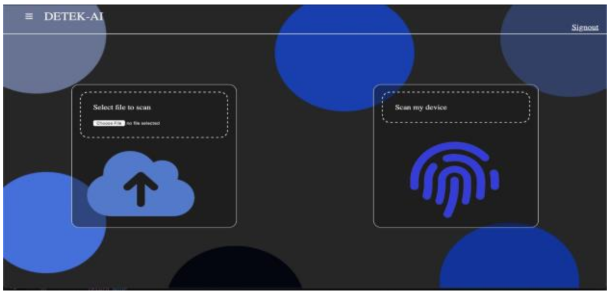
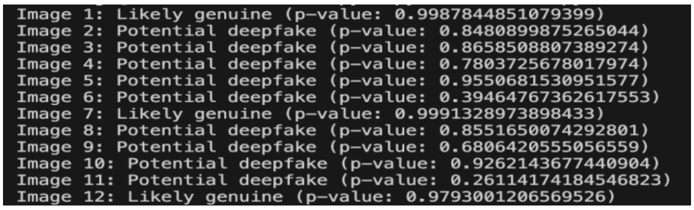


Fig3.2 services of DETEK-AI



```
Image 1: Likely genuine (p-value: 0.9987844851079399)
Image 2: Potential deepfake (p-value: 0.8480899875265044)
Image 3: Potential deepfake (p-value: 0.8658508807389274)
Image 4: Potential deepfake (p-value: 0.7803725678017974)
Image 5: Potential deepfake (p-value: 0.9550681530951577)
Image 6: Potential deepfake (p-value: 0.39464767362617553)
Image 7: Likely genuine (p-value: 0.9991328973898433)
Image 8: Potential deepfake (p-value: 0.8551650074292801)
Image 9: Potential deepfake (p-value: 0.6806420555056559)
Image 10: Potential deepfake (p-value: 0.9262143677440904)
Image 11: Potential deepfake (p-value: 0.26114174184546823)
Image 12: Likely genuine (p-value: 0.9793001206569526)
```

Fig3.3 scan_results.log generated

## IV. METHODOLOGY

- **Libraries:**
  - **a. os:** This module offers an approach to use operating system-dependent functionality that is portable.
  - **b. cv2:** For image processing jobs, OpenCV (Open Source Computer Vision Library) is utilized.
  - **c. numpy:** A library for numerical computing used to manipulate arrays.
  - **d. SciPy.stats.chisquare:** A tool for doing the goodness of fit chi-square test.

```
def benfords_law_test(data):
    observed_counts =
[data.count(str(digit)) for digit in
range(1, 10)]
…
```

- A list of data is the input for this function.
- It determines the observed counts of the data's leading digits, which range from 1 to 9.
- computes the anticipated counts using Benford's Law.
- computes the observed and expected frequencies and normalizes the counts.
- utilizes the chi-square test to compare the frequencies that are seen and predicted.
- gives back the p-value that was determined by the chi-square test.

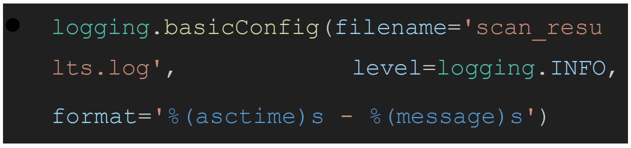
```
def analyze_image(image_path):
    img = cv2.imread(image_path,
cv2.IMREAD_GRAYSCALE)
```

- This method uses OpenCV in grayscale mode to interpret an image.
- In the event that the image loads successfully, pixel values are converted to strings and the image is flattened into a 1D array.
- gives back the plowed picture in features format.

```
def main(image_folder):
    p_values = []
…
```

- Through every image file in the designated folder, this function iterates.

- uses the analyze_image function to get each image's characteristics.
- uses the benfords_law_test function to analyze the retrieved characteristics using Benford's Law.
- logs each image's p-value calculation.
- identifies photos as maybe authentic or possibly deepfake based on their p-values.



```
logging.basicConfig(filename='scan_results.log', level=logging.INFO, format='%(asctime)s - %(message)s')
```

- configures the logging system to write log messages to the scan_results.log file.

These algorithms and techniques collectively contribute to a comprehensive approach in detecting deepfake, by analyzing both visual features and underlying data. By leveraging the power of Benford's Law, we enhance the accuracy and effectiveness of our deepfake detection system.

## V. KEY FEATURES

- **Benford's Law Application:** Benford's Law has been used in various fields for anomaly detection, making it a unique approach for identifying potential irregularities in image datasets.
- **Threshold Setting:** By setting a threshold, the code allows for customization based on the specific use case, enabling users to adjust sensitivity and reduce false positives or negatives.
- **Authentication of Digital Image:** Verifying the authenticity of digital images by analyzing the distribution of leading digits can be crucial in contexts like journalism, forensics, and legal proceedings.
- **Ensuring Data Integrity:** This algorithm ensures that no data is collected as the user data is converted into integer numbers for deepfake detection.
- **Rapid Computation**:This algorithm converts the data into integer values for deepfake detection, enhancing the computation speed.

## V. SOFTWARE & HARDWARE REQUIREMENTS

- **Software**
  1. **Front-end framework:** (HTML, CSS) User-friendly interface
  2. **Database:** a database system (PHP) to store user data or other relevant information.
  3. **Python Library:** Used for implementation of logic.
  4. **Image/Video Processing Library:** A library for processing images and videos. OpenCV is commonly used for this purpose.
- **Hardware**
  1. **Memory (RAM):** Minimum 4GB RAM to handle the algorithm inference and concurrent user requests.
  2. **OS:** Windows, MacOs, Linux.

## VI. CONCLUSION

To sum up, the examined research publications stress the vital significance of deepfake detection and offer a wide range of approaches to address the problem. These tactics cover a wide range of techniques, from cutting-edge ones like deep-CNN architectures, blockchain technology, and cyber vaccine to more conventional ones like text-based detection with deep learning models and word embeddings. High accuracy scores are attained in identifying deepfakes in a variety of media types using all methods.

But in order to detect deepfakes accurately, a number of things need to be carefully taken into account, including feature extraction, model selection, and dataset development. The effectiveness of each technique may differ based on the particular deepfake type and dataset used. However, these articles offer a variety of useful methods for recognizing and detecting fake information, which is a significant advancement in the ongoing fight against the spread of deepfakes.

By introducing a fresh technique to deepfake detection, the research presents the DETEK-AI system, which builds upon this body of knowledge. Utilizing Benford's Law in conjunction with an adjustable classification threshold, this system provides a reliable way to verify digital photographs and guarantee data integrity. DETEK-AI has potential for use in forensics, judicial proceedings, journalism, and other sectors because of its quick analysis of picture datasets and capacity to spot possible deepfakes. All things considered, the research articles' and the DETEK-AI system's contributions constitute noteworthy advancements in tackling the problems caused by deepfake technology.

## VII. REFERENCES

[1] Li, X., Lyu, S., Zhang, H., & Kewei, S. (2020). Deepfake detection based on fusion network with adaptive feature coupling. Neurocomputing, 387, 64-73.

[2] Wang, X., Ling, G., Song, X., Zhang Y., & Zhang, L. (2021). Survey of Deepfake Forensics and Detection Techniques. IEEE Access, 9, 16110-16127.

[3] S. Sadiq, T. Aljrees and S. Ullah, "Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets," in IEEE Access, vol. 11, pp. 95008-95021, 2023, doi: 10.1109/ACCESS.2023.3308515.

[4] C. -C. Chang, H. H. Nguyen, J. Yamagishi and I. Echizen, "Cyber Vaccine for Deepfake Immunity,"

in IEEE Access, vol. 11, pp. 105027-105039, 2023, doi: 10.1109/ACCESS.2023.3311461.

[5] Y. Patel et al., "An Improved Dense CNN Architecture for Deepfake Image Detection," in IEEE Access, vol. 11, pp. 22081-22095, 2023, doi: 10.1109/ACCESS.2023.3251417.

[6] N. M. Alnaim, Z. M. Almutairi, M. S. Alsuwat, H. H. Alalawi, A. Alshobaili and F. S. Alenezi, "DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms," in IEEE Access, vol. 11, pp. 16711-16722, 2023, doi: 10.1109/ACCESS.2023.3246661.

[7] J. A. Costales, S. Shiromani and M. Devaraj, "The Impact of Blockchain Technology to Protect Image and Video Integrity from Identity Theft using Deepfake Analyzer," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 730-733, doi: 10.1109/ICIDCA56705.2023.10099668.

[8] A. Das, K. S. A. Viji and L. Sebastian, "A Survey on Deepfake Video Detection Techniques Using Deep Learning," 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 2022, pp. 1-4, doi: 10.1109/ICNGIS54955.2022.10079802.

[9] N. Khatri, V. Borar and R. Garg, "A Comparative Study: Deepfake Detection Using Deep-learning," 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2023, pp. 1-5, doi: 10.1109/Confluence56041.2023.10048888.

[10] A. H. Khalifa, N. A. Zaher, A. S. Abdallah and M. W. Fakhr, "Convolutional Neural Network Based on Diverse Gabor Filters for Deepfake Recognition," in IEEE Access, vol. 10, pp. 22678-22686, 2022, doi: 10.1109/ACCESS.2022.3152029.

[11] Y. -X. Luo and J. -L. Chen, "Dual Attention Network Approaches to Face Forgery Video Detection," in IEEE Access, vol. 10, pp. 110754-110760, 2022, doi: 10.1109/ACCESS.2022.3215963.

[12] W. Shahid, Y. Li, D. Staples, G. Amin, S. Hakak and A. Ghorbani, "Are You a Cyborg, Bot or Human?—A Survey on Detecting Fake News Spreaders," in IEEE Access, vol. 10, pp. 27069-27083, 2022, doi: 10.1109/ACCESS.2022.3157724.

[13] V. -N. Tran, S. -G. Kwon, S. -H. Lee, H. -S. Le and K. -R. Kwon, "Generalization of Forgery Detection With Meta Deepfake Detection Model," in IEEE Access, vol. 11, pp. 535-546, 2023, doi: 10.1109/ACCESS.2022.3232290.

[14] A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 18757-18775, 2022, doi: 10.1109/ACCESS.2022.3151186.

[15] T. Dar, A. Javed, S. Bourouis, H. S. Hussein and H. Alshazly, "Efficient-SwishNet Based System for Facial Emotion Recognition," in IEEE Access, vol. 10, pp. 71311-71328, 2022, doi: 10.1109/ACCESS.2022.3188730.

[16] J. Kang, S. -K. Ji, S. Lee, D. Jang and J. -U. Hou, "Detection Enhancement for Various Deepfake Types Based on Residual Noise and Manipulation Traces," in IEEE Access, vol. 10, pp. 69031-69040, 2022, doi: 10.1109/ACCESS.2022.3185121.

[17] M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 25494-25513, 2022, doi: 10.1109/ACCESS.2022.3154404.

[18] D. B. Frolov, D. D. Makhaev and V. V. Shishkarev, "Deepfakes and Information Security Issues," 2022 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS), Saint Petersburg, Russian Federation, 2022, pp. 147-150, doi: 10.1109/ITQMIS56172.2022.9976507.

[19] H. Ilyas, A. Irtaza, A. Javed and K. M. Malik, "Deepfakes Examiner: An End-to-End Deep Learning Model for Deepfakes Videos Detection," 2022 16th International Conference on Open Source Systems and Technologies (ICOSST), Lahore, Pakistan, 2022, pp. 1-6, doi: 10.1109/ICOSST57195.2022.10016871.

[20] J. John and B. V. Sherif, "Comparative Analysis on Different DeepFake Detection Methods and Semi Supervised GAN Architecture for DeepFake Detection," 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Dharan, Nepal, 2022, pp. 516-521, doi: 10.1109/I-SMAC55078.2022.9987265.

[21] K. Jalui, A. Jagtap, S. Sharma, G. Mary, R. Fernandes and M. Kolhekar, "Synthetic Content

Detection in Deepfake Video using Deep Learning," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2022, pp. 01-05, doi: 10.1109/GCAT55367.2022.9972081.

[22] S. S. Chauhan, N. Jain, S. C. Pandey and A. Chabaque, "Deepfake Detection in Videos and Picture: Analysis of Deep Learning Models and Dataset," 2022 IEEE International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2022, pp. 1-5, doi: 10.1109/ICDSIS55133.2022.9915885.

[23] V. Jolly, M. Telrandhe, A. Kasat, A. Shitole and K. Gawande, "CNN based Deep Learning model for Deepfake Detection," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-5, doi: 10.1109/ASIANCON55314.2022.9908862.

[24] A. Rahman et al., "Short And Low Resolution Deepfake Video Detection Using CNN," 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), Hyderabad, India, 2022, pp. 259-264, doi: 10.1109/R10-HTC54060.2022.9929719.

[25] S. Yadav, S. Bommareddy and D. K. Vishwakarma, "Robust and Generalized DeepFake Detection," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984553