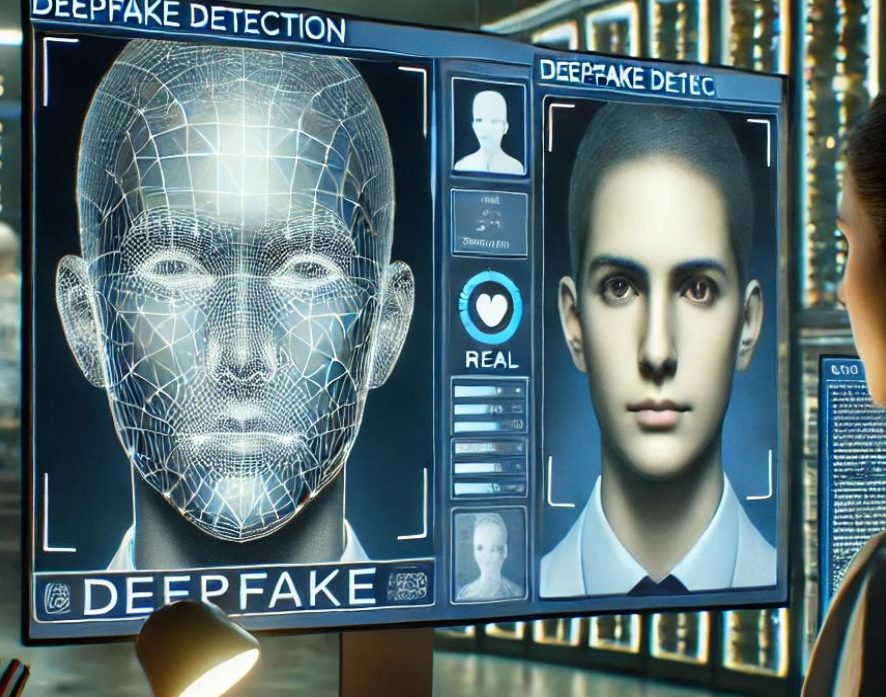
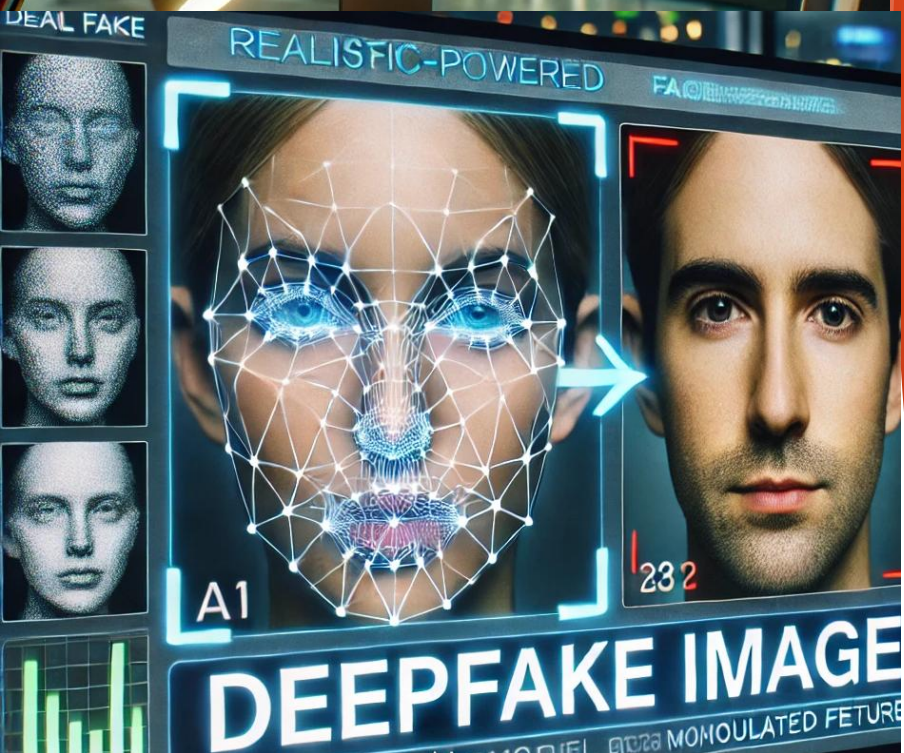




Title: Sustainable AI: Assessing Performance in Detecting Fake Images



BY NAGA SATYA VARMA CHAMPATI

ID:2410049

Introduction



Detection of Deepfakes:

Deepfake images are artificially generated counterfeit photographs employed for the dissemination of misinformation.

Artificial intelligence models such as Convolutional Neural Networks and Transformers facilitate the identification of counterfeit photographs.

However, AI models require excessive energy, rendering detection unsustainable.

The significance of sustainable AI.

Artificial intelligence requires substantial electricity for training and operation.

Elevated energy use leads to increased expenses and a substantial carbon footprint.

We require an equilibrium between precision and energy efficiency.

Research Problem

AI models effectively identify deepfakes; yet they consume excessive energy.

Real-time detection is inefficient and costly.

Most studies prioritize accuracy while neglecting energy use.

No definitive AI model achieves an equilibrium between performance and efficiency.

Research Aim & Objectives

Aim : To assess how AI-based **fake image detection models** can be made **sustainable**.

Objectives :

1. To analyze the computational expense of AI deepfake detection algorithms.
2. To ascertain obstacles in the real-time processing of deepfake images.
3. Investigate low-energy artificial intelligence models that preserve elevated detection precision.

Literature Review

Deepfake detection has emerged as a vital study domain owing to the swift proliferation of AI-generated forgeries.

visual representations. A multitude of scholars have concentrated on Convolutional Neural Networks (CNNs) and Transformers.

Model-based approaches for the detection of counterfeit photographs. Although these models attain elevated accuracy, they necessitate

substantial computational resources, rendering them impractical for real-world applications.

- Numerous methodologies have been suggested for deepfake detection, classified into three categories.

Literature Review (cont.)

Current AI-Driven Techniques for Detecting Fabricated Images Various methodologies have been suggested for the detection of deepfakes:

CNN-Based Methodologies

VGGNet and ResNet are extensively utilized for deepfake detection [1].
Convolutional Neural Networks excel on particular datasets but encounter difficulties with unfamiliar deepfakes [2].
Significant computing expense and restricted generalization are primary challenges.

Transformer-Based Models

Vision Transformers (ViTs) surpass Convolutional Neural Networks (CNNs) in feature extraction.
Demand extensive datasets and substantial computational resources

Literature Review (cont.)

Hybrid AI Models (Convolutional Neural Networks + Transformers)

- Hybrid models integrate the feature extraction capabilities of CNNs with the contextual learning of Transformers.
- More precise and broadly applicable, albeit computationally intensive [4].

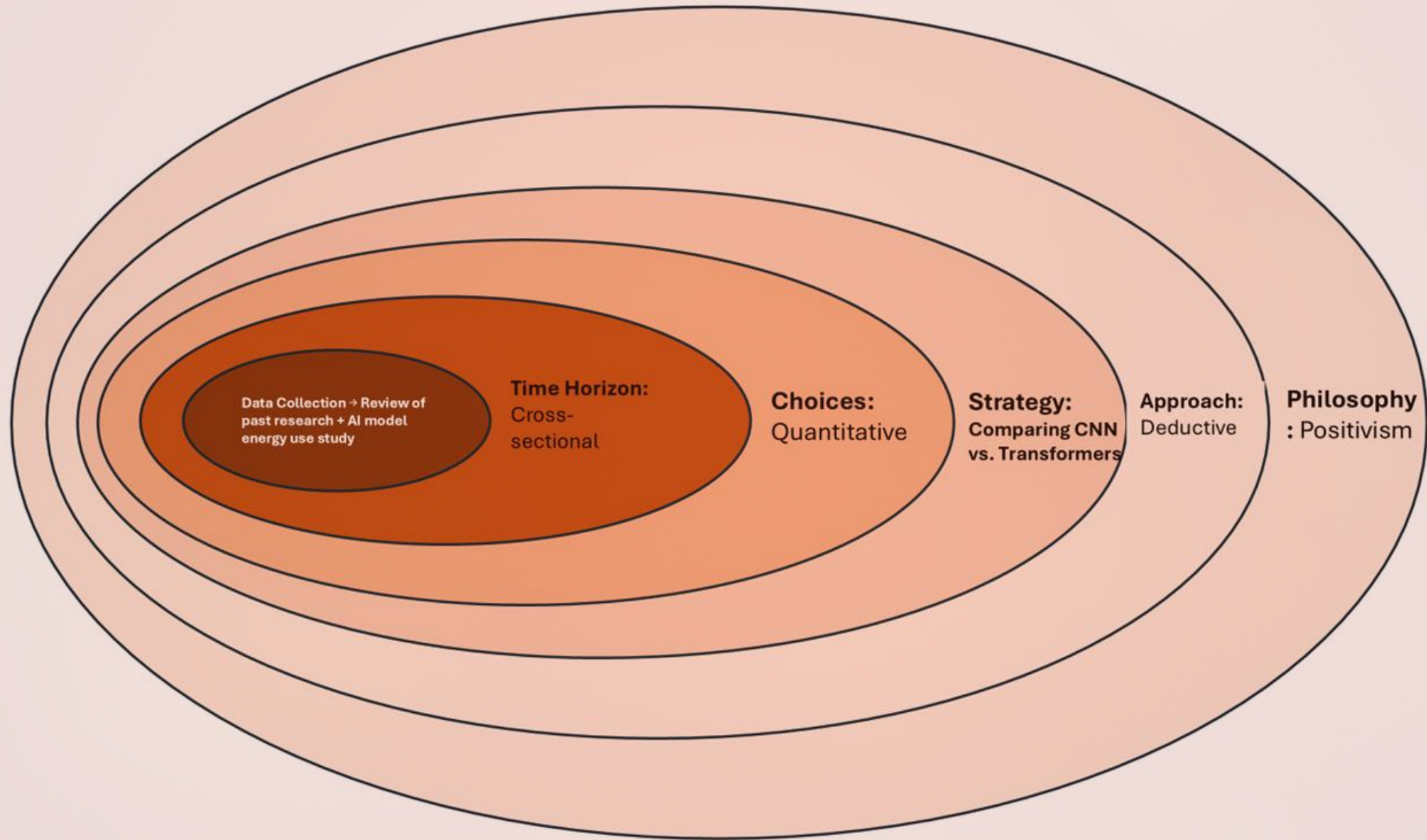
Challenges in AI-Driven Deepfake Identification

- The substantial processing expense renders real-time detection challenging [5].
- AI models perform effectively solely on familiar datasets but struggle with novel deepfake variants [6].
- There is a lack of significant study on the energy efficiency of AI in deepfake detection.

Literature Review (cont.)

AI Model	Advantages	Disadvantages
CNN-Based Models	High accuracy on known datasets.	Struggles with new deepfake types.
Transformer Models	Strong contextual learning, handles complex patterns.	Computationally expensive, needs massive data.
Hybrid (CNN + Transformer)	Balances CNN's feature extraction with Transformer's learning.	Requires large-scale training, expensive.

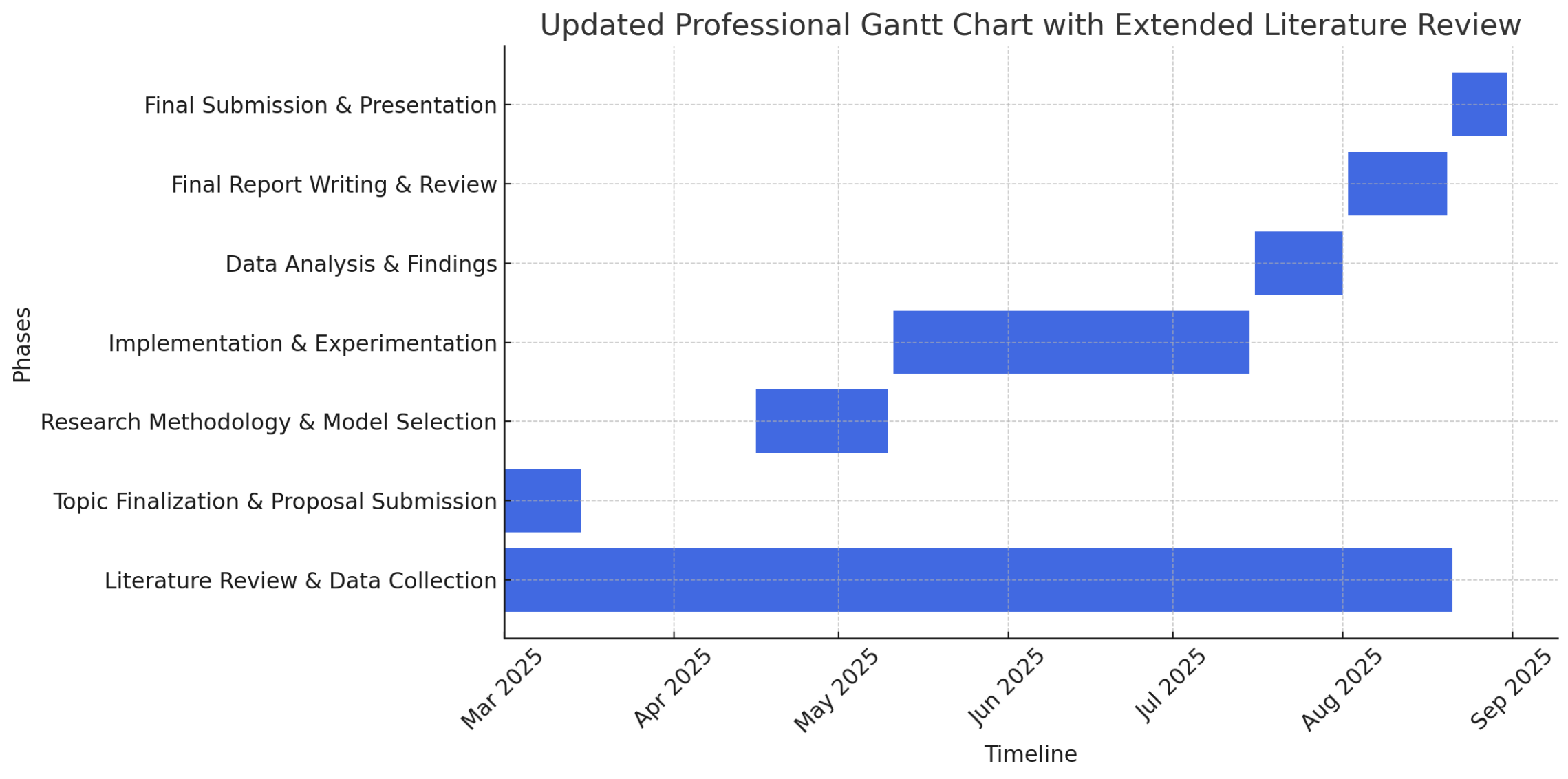
Onion Research Framework



Research Methodology

- 1) Philosophy (Positivism): Prioritizes objective, quantifiable data for the evaluation of hypotheses.
- 2) Methodology (Deductive): Evaluates pre-established hypotheses through targeted research inquiries.
- 3) Strategy (Comparative Analysis): Assesses performance disparities between CNNs and Transformers.
- 4) Options (Quantitative): Depends on numerical measures (accuracy, F1 score) for model evaluation.
- 5) Time Horizon (Cross-sectional): Evaluates data at a singular moment, emphasizing current model representations.
- 6) Data Collection (Literature + Empirical): Integrates existing research reviews and energy consumption indicators for model assessment.

Project Planning and Resources:



Gantt Chart

Project Impact & Limitations

Impact of the study:

This study emphasizes the significance of reconciling energy efficiency with precision in artificial intelligence models.

It can assist AI developers in formulating sustainable deepfake detection algorithms that operate effectively in real-world environments.

Policymakers and companies can leverage these findings to save computing expenses and formulate AI plans with reduced carbon emissions.

Limitations of the study:

The research depends on pre-existing AI models and datasets, and real-time application has not been evaluated.

Constraints on computational resources may restrict the evaluation of all available models.

Future enhancements require practical implementation and extensive performance evaluation throughout time.

Conclusion & Future Work

Conclusion:

AI-driven deepfake detection is remarkably efficient yet resource-intensive.

Sustainable AI must prioritize minimizing energy consumption while preserving precision.

Hybrid models (CNN + Transformers) present a viable approach but necessitate more improvement.

Prospective Endeavors:

Creating lightweight deepfake detection models that operate efficiently on edge devices.

Deploying real-time deepfake detection algorithms with reduced energy expenditure.

Additional investigation on trimming and quantization methodologies to improve energy efficiency.

References (IEEE Format)

- [1] K. Korshunov and S. Marcel, "Deepfake detection: CNN-based approaches," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 1, pp. 18-30, 2022.
- [2] A. Tolosana, R. Vera-Rodriguez, J. Fierrez, M. Morales, and J. Ortega-Garcia, "DeepFakes and beyond: A survey of face manipulation and fake detection," *Information Fusion*, vol. 64, pp. 131-148, 2020.
- [3] T. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [4] J. Yu et al., "FaceForensics++: Learning to detect manipulated facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 203-220, 2021.
- [5] R. Zhang, "Deepfake detection: Challenges and future directions," *Journal of Artificial Intelligence Research*, vol. 65, pp. 265-280, 2022.
- [6] M. Wang et al., "Generalization of AI-based deepfake detection across datasets," *IEEE Transactions on Image Processing*, vol. 30, pp. 232-245, 2021.



Thank You !!!

-Naga Satya Varma Champati