AI-Driven Distributed Data Management in Deepfake Images in Smart Cities

1 Theory Plan

In the modern era, deepfake images pose significant threats to smart city infrastructure, ranging from misinformation to security risks. Managing and detecting deepfakes require an AI-driven distributed approach that ensures secure data handling while maintaining privacy and authenticity. This research aims to explore deepfake detection methods, blockchain-based data management, and AI-driven solutions that can be implemented in a decentralized manner.

The theoretical foundation of this study will involve an extensive literature review, focusing on deep learning models such as Convolutional Neural Networks (CNNs), Transformer-based architectures, and hybrid detection techniques. Additionally, blockchain will be analyzed as a tool for ensuring data integrity by implementing smart contracts for image authentication. The study will also investigate federated learning as a method for training models on distributed networks without exposing sensitive data. Existing datasets, such as the DeepFake Detection Challenge and FaceForensics++, will be reviewed for benchmarking AI models.

Furthermore, challenges such as adversarial attacks, computational overhead, and ethical concerns in deploying deepfake detection in smart cities will be addressed. This research will provide insights into future developments, highlighting the role of AI and decentralized networks in mitigating deepfake-related threats in urban environments.

2 Lab Plan

The practical implementation will focus on developing a basic deepfake detection model using machine learning techniques. The primary objective is to train and evaluate a classification model that can distinguish between real and deepfake images. Python will be the primary programming language, utilizing libraries such as TensorFlow, OpenCV, and DeepFace for facial analysis.

The first step involves collecting a dataset, such as a subset of the FaceForensics++ or DeepFake Detection Challenge dataset. Preprocessing will include resizing images, grayscale conversion, and normalization to ensure consistency. Facial feature extraction will be conducted using OpenCV's face detection or DeepFace's embedding techniques. A Convolutional Neural Network (CNN), such as MobileNet or VGG16, will be fine-tuned to classify images as real or fake. The model will be trained on a small dataset to allow efficient computation and quick performance evaluation.

After training, performance will be assessed using metrics such as accuracy, precision, recall, and confusion matrices. Additionally, the model's effectiveness will be tested on unseen images to analyze generalization capability. If computational resources allow, an optional implementation step will include deploying the trained model on an edge device like a Raspberry Pi to simulate real-world application scenarios.

The expected outcome of this lab work is a functional deepfake detection system that demonstrates AI's capability in identifying manipulated images. Limitations such as dataset size, training time, and hardware constraints will be discussed, along with potential future improvements such as GAN-based detection models and blockchain integration for enhanced security.