DEEPFAKE CONTENT TYPES AND THEIR GENERATION METHODS

Primbetov Abbaz[1], Normo‘minov Anvarjon[2]

[1]Phd student, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi
[1,2]Senior lecturer, Computer engineering, Tashkent University of Applied Sciences
abbaz0203@mail.ru, anormuminov072@gmail.com.

Annotation: Deepfake technology, powered by artificial intelligence, enables the manipulation of media content in multiple ways, including identity swapping, facial expression alteration, attribute modification, background replacement, and realistic speech or image synthesis. These manipulations can create highly convincing yet artificial content, posing challenges for digital media authentication, forensic analysis, and cybersecurity. Understanding the various deepfake types and their generation techniques is essential for designing robust detection methods and improving the reliability of automated verification systems.

Keywords: Deepfake, Artificial Intelligence, Identity Swapping, Expression Manipulation, Facial Attribute Editing, TTS, Image/Video Synthesis.

INTRODUCTION

he rapid advancement of artificial intelligence (AI) and deep generative models has revolutionized the creation and manipulation of digital media. Among these, deepfake technology—which enables the realistic synthesis or alteration of images, videos, and audio—has become a significant concern for media integrity, cybersecurity, and forensic science. Deepfakes can convincingly modify facial identities, expressions, attributes, backgrounds, and even generate synthetic speech, making the detection of manipulated content increasingly challenging.

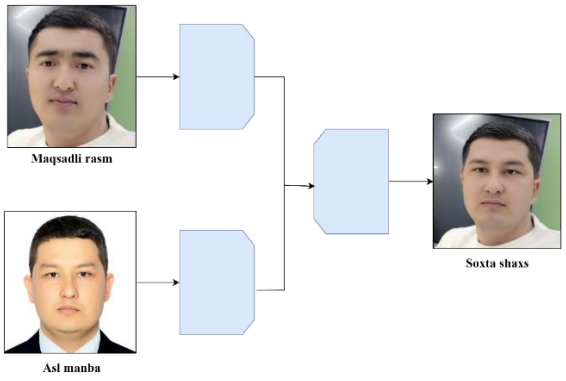
The growing availability of deepfake generation tools and large-scale datasets has democratized the production of highly realistic yet synthetic content, raising serious ethical, social, and legal implications. In forensic and security contexts, accurate identification of deepfakes is essential to prevent misinformation, protect individual identity, and ensure the credibility of evidence.

This study focuses on the systematic categorization of deepfake types, the underlying generation techniques, and the development of robust detection methods. By analyzing identity swapping, expression manipulation, facial attribute editing, background replacement, and multimodal synthesis, this work aims to enhance the reliability of automated forensic tools and contribute to the ongoing efforts in combating digital media manipulation.

METHODOLOGY

Deepfake content is generated using various artificial intelligence (AI) approaches that manipulate specific aspects of a media file, such as identity, expression, background, text, or audio. These manipulations produce highly realistic synthetic content while retaining certain aspects of the original media. In this study, we focus on seven major types of deepfake content, each associated with distinct AI-based generation techniques. Representative illustrations for each type are provided to visually demonstrate the manipulation process.

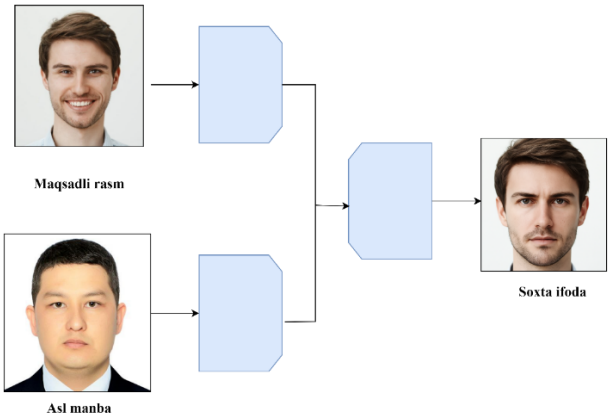
1. Identity Swapping (Face/Voice Substitution): Identity swapping involves transferring the facial or vocal characteristics of a source person onto a target person. In images and video, this is commonly referred to as face swapping, while in audio, it is known as voice conversion. The primary goal is to alter the target’s facial expressions or voice timbre to create realistic yet unidentifiable content.



(a) Shaxs yuzini almashtirish

Figure 1: Example of identity swapping with source and target faces/voices.

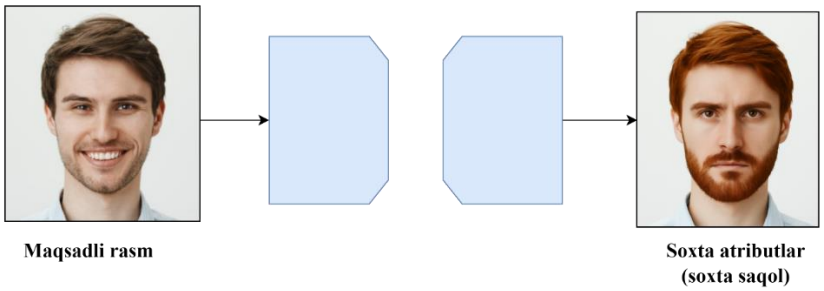
2. Emotion/Expression Swapping: In emotion or expression swapping, the person's identity remains unchanged; only facial expressions or vocal emotions are transformed. Techniques such as face reenactment are used to replicate facial movements or lip synchronization, enabling realistic animation of expressions without changing the underlying identity.



***(b)*Emosiyalarni almashtirish**

Figure 2: Example of facial expression reenactment.

3. Facial Attribute Manipulation: Facial attribute manipulation modifies age, gender, skin tone, hairstyle, or other facial features while preserving the individual's identity. These generative algorithms are often used for cosmetic editing, digital makeup, or controlled changes in facial appearance.

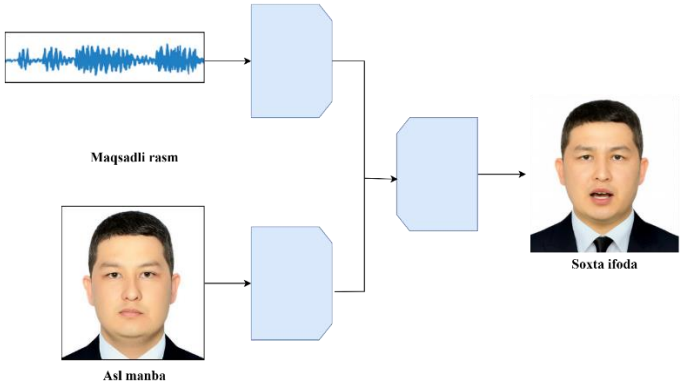


(c) Yuz atributlarini o‘zgartirish

Figure 3: Examples of facial attribute modification—age, gender, hairstyle.

4. Talking Face Synthesis: Talking face synthesis generates realistic lip, head, and facial movements based on text, audio, or multimodal input. This approach enables the creation of

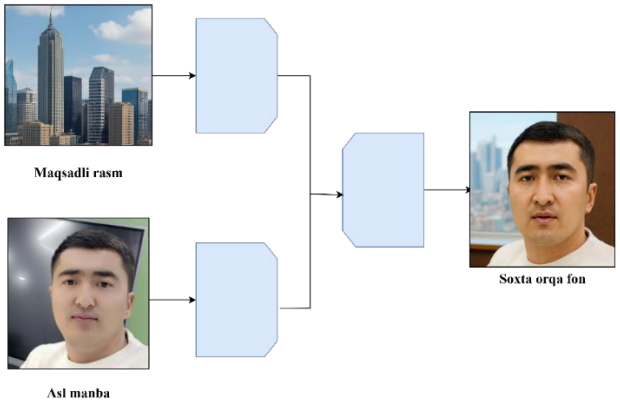
synchronized talking avatars or video content, preserving natural motion dynamics for more convincing results.



(d) **Gapirayotgan yuzlarni yaratish**

Figure 4: Example of talking face synthesis from audio input.

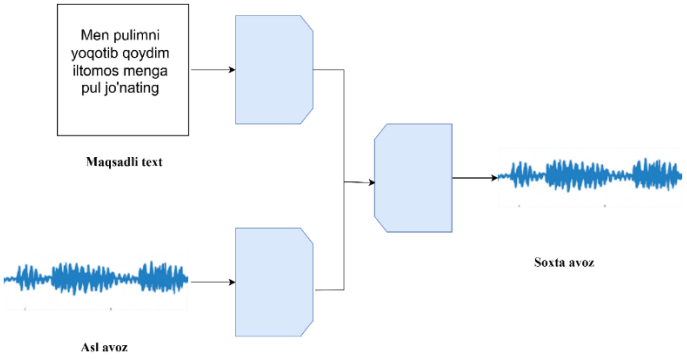
5. Background Swapping: Background swapping maintains the subject while altering the surrounding scene or audio environment. In images and video, this involves segmenting the foreground and placing it into a new background, whereas in audio, ambient noise or environmental sounds can be replaced.



(F) Fonni almashtirish

Figure 5: Example of background replacement in video/image.

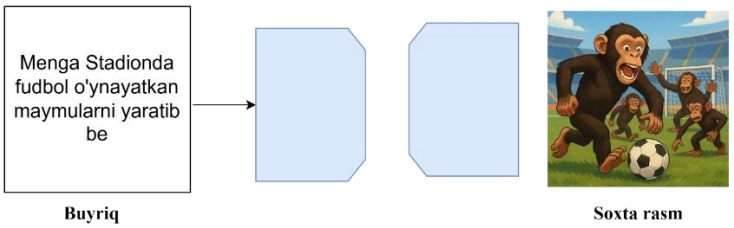
6. Text-to-Speech (TTS) Synthesis: Text-to-speech synthesis converts written text into natural-sounding speech. Advanced TTS models, such as ElevenLabs, can replicate the voice characteristics of specific individuals, producing speech that is highly realistic and often indistinguishable from genuine recordings.



(j) Matndan nutq hosil qilish

Figure 6: TTS waveform and spectrogram for generated speech.

7. Text-to-Image/Video Generation: Text-to-image and text-to-video models, including Stable Diffusion, DALL-E, and GLIDE, generate entirely synthetic visual content from textual prompts. These models are capable of producing high-quality images and videos that correspond closely to descriptive input, enabling flexible and creative media synthesis.



(k) Matndan tasvir/video hosil qilish

Figure 7: Example of a text-to-image/video generation prompt and output.

The above deepfake types can be domain-specific (e.g., image, audio, video, text) or domain-agnostic, depending on the targeted media. Robust detection algorithms must be tailored to the characteristics of each type to reliably identify manipulated content. The combination of spatial, temporal, and multimodal features is critical for effective deepfake detection.

CONCLUSION

The rapid advancement of deep generative models has significantly increased the realism and accessibility of deepfake content, posing challenges for media authentication, cybersecurity, and forensic investigations. This study has systematically examined the major types of deepfake content, including identity swapping, emotion and expression manipulation, facial attribute editing, talking face synthesis, background replacement, text-to-speech synthesis, and text-to-image/video generation. Each type utilizes distinct AI-based techniques that can produce convincing yet artificial media, emphasizing the need for specialized detection methods.

By understanding the characteristics and generation methods of different deepfake types, researchers and practitioners can develop robust detection algorithms capable of addressing both domain-specific and domain-agnostic manipulations. Hybrid approaches combining spatial, temporal, and multimodal analysis demonstrate significant potential in improving detection accuracy.

Ultimately, this work highlights the importance of continuous research in deepfake detection and forensic analysis to safeguard digital media integrity, prevent misinformation, and enhance trust in digital content. Future work may focus on integrating real-time detection systems, cross-modal verification, and the development of standardized benchmarks to further strengthen the reliability of deepfake identification.

REFERENCE:

1. M. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “Deepfakes and beyond: A survey of face manipulation and fake detection,” *Information Fusion*, vol. 64, pp. 131–148, 2020.
2. Diel et al., “Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers,” *Computers in Human Behavior Reports*, vol. 16, p. 100538, 2024.
3. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251–1258.
4. M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2019, pp. 6105–6114.
5. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1492–1500.

6. W. van Gansbeke, A. Dhoedt, and J. Van de Weijer, “Temporal awareness in deepfake detection,” *IEEE Trans. Biometrics, Behavior, and Identity Science*, vol. 3, no. 2, pp. 176–187, 2021.
7. S. Tipper, H. F. Atlam, and H. S. Lallie, “An investigation into the utilisation of CNN with LSTM for video deepfake detection,” *Applied Sciences*, vol. 14, no. 21, p. 9754, 2024.
8. Maxmudjanov Sarvar, Primbetov Abbaz Muratbay Uli, and Naimov Axadjon Tojimirza O‘G‘Li. "DEEPFAKE DETECTION USING A HYBRID RESNEXT AND LSTM ARCHITECTURE." Al-Farg’oniy avlodlari 1.2 (2025): 87-94.