
# A Look at Open-Source Deepfake Detection

Tattle

Priyash Shah, Aatman Vaidya

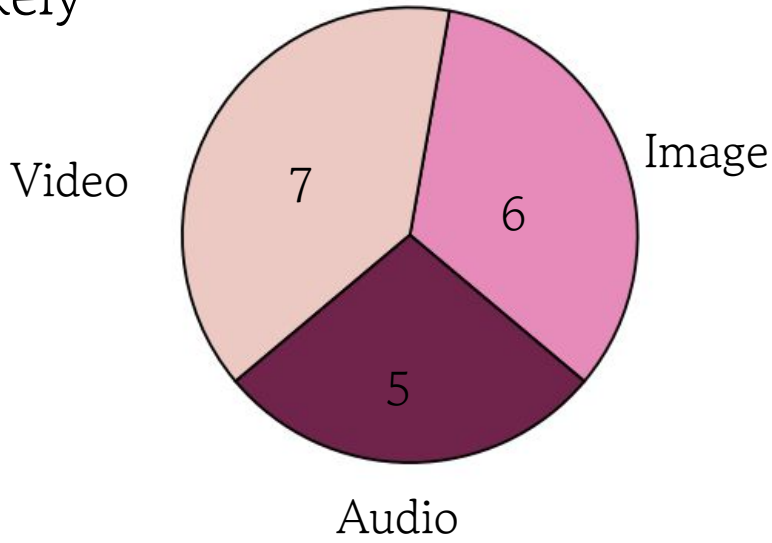
# Motivation

- Fact-checkers and organizations use a range of automated AI tools, both proprietary and open-source, to identify deepfakes and media manipulation in video/audio.
- We wanted to gain a better understanding of the performance of open-source tools and machine learning models.
- Ease of self hosting
- Tradeoff between proprietary and open-source tools

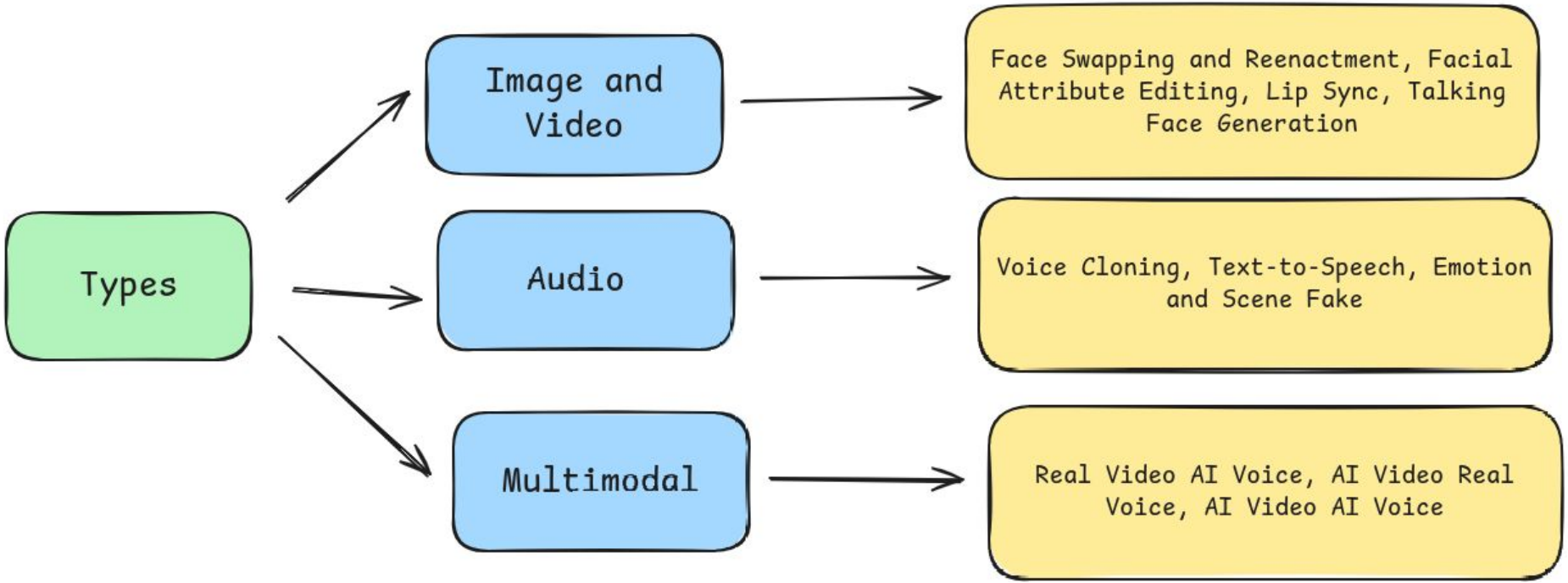
# Deepfake-o-Meter

DeepFake-O-Meter

- 18 state-of-the-art deepfake detection ML models.
- It is an active platform, and more likely to have the best-performing methods and models.



# Types


Types
Image and Video
Audio
Multimodal
Face Swapping and Reenactment, Facial Attribute Editing, Lip Sync, Talking Face Generation
Voice Cloning, Text-to-Speech, Emotion and Scene Fake
Real Video AI Voice, AI Video Real Voice, AI Video AI Voice

# Methods

Face Swapping and Reenactment, Facial Attribute Editing, Lip Sync, Talking Face Generation

→

CNN-based, Global image context and local patch features, GAN-based, Vision Transformer based, Neighbouring Pixel Relationships (NPR), Spatial(frame-level) and Temporal(video-level) attention feature extraction, Temporal transformer, Latent Space Data Augmentation (LSDA), spatial-temporal inconsistency extractor

Voice Cloning, Text-to-Speech, Emotion and Scene Fake

→

Linear Frequency Cepstral Coefficient (LFCC) classification, Feature extractor with a vocoder identification module, Whisper encoder as a feature extractor, Spectrogram waveform classification

# Examples

## RawNet2-Vocoder

Is Synthetically Generated?

Arvind Kejriwal   

Narendra Modi   

voice clone example, generated by using Coqui's XTTS-V2 voice cloning TTS model - https://huggingface.co/coqui/XTTS-v2

# Examples

**Whisper** (finetuned ensemble)

# Examples

**LSDA** (Latent Space Data Augmentation)

Is Synthetically Generated?

20% likelihood

Audio Generated using Indic Parler TTS -

# Examples

## LSDA (Latent Space Data Augmentation)


Is Synthetically Generated?

40% likelihood

Audio Generated using Indic Parler TTS -

# Some Takeaways

- Most image and video models only support detection of facial attributes based manipulation

# Some Takeaways

- Most image and video models only support detection of facial attributes based manipulation

- Models still face challenges in detecting low-resolution manipulations or previously unseen real-world images/audio.

# Some Takeaways

- Most image and video models only support detection of facial attributes based manipulation
- Models still face challenges in detecting low-resolution manipulations or previously unseen real-world images/audio.
- Audio-based models are trained on good quality controlled datasets and failed to perform well on real world examples.

# Some Takeaways

- Most image and video models only support detection of facial attributes based manipulation
- Models still face challenges in detecting low-resolution manipulations or previously unseen real-world images/audio.
- Audio-based models are trained on good quality controlled datasets and failed to perform well on real world examples.
- Open source model have visibility into training data and architecture.

# Scan to read our Blog

Tattle