

The DeepSpeak Dataset

Sarah Barrington¹, Maty Bohacek², Hany Farid¹

¹University of California, Berkeley, ²Stanford University

Abstract

Deepfakes represent a growing concern across domains such as disinformation, fraud, and non-consensual media. In particular, the rise of video conference and identity-driven attacks in high-stakes scenarios—such as impostor hiring—demands new forensic resources. Despite significant efforts to develop robust detection classifiers to distinguish the real from the fake, commonly used training datasets remain inadequate: relying on low-quality and outdated deepfake generators, consisting of content scraped from online repositories without participant consent, lacking in multimodal coverage, and rarely employing identity-matching protocols to ensure realistic fakes. To overcome these limitations, we present the DeepSpeak dataset, a diverse and multimodal dataset comprising over 100 hours of authentic and deepfake audio-visual content, specifically focused on the challenging and diverse “talking heads” context. We contribute: i) more than 50 hours of real, self-recorded data collected from 500 diverse and consenting participants, ii) more than 50 hours of state-of-the-art audio and visual deepfakes generated using 14 video synthesis engines and three voice cloning engines, and iii) an embedding-based, identity-matching approach to ensure the creation of convincing, high-quality identity face swaps that realistically simulate adversarial deepfake attacks. We also perform large-scale evaluations of state-of-the-art deepfake detectors and show that, without retraining, these detectors fail to generalize to this DeepSpeak dataset, highlighting the importance of a large and diverse dataset containing deepfakes from the latest generative-AI tools.

<https://github.com/hfaridlab/deepspeak>

1. Introduction

Today, generative-AI is capable of creating hyper-realistic images [36], voices [4], and videos [19] of people talking or doing just about anything. These technologies hold the promise to both revolutionize many industries while also amplifying the spread and belief in dangerous lies and conspiracies [10, 49], interfering with elections [16, 45], supercharging small- and large-scale fraud [5], and – seemingly

unable to escape its roots – continue to be used in the creation of non-consensual intimate imagery (NCII) [13, 50].

Scalable, generalizable, and accurate detection of deepfakes has, therefore, become a pressing problem with deep social, political, and economic implications. At the same time, the nascent digital-forensics community has struggled with the lack of large-scale, high-quality, up-to-date, and ethically collected datasets for training and evaluation.

In this work, we introduce an audio and video dataset designed to aid the digital-forensic, computer-vision, and broader AI-safety communities. This dataset consists of 100 hours of real and deepfake video of people talking and gesturing. The real videos were self-recorded with consent from the participants using their own hardware, ensuring a wide range of recording environments, hardware variability and identities, a crucial component for the development of robust detectors. These deepfakes consist of avatar deepfakes (from three generators), face-swap deepfakes (with multiple variants from three generators), lip-sync deepfakes (from four generators), and audio deepfakes (from three generators) spliced into a subset of the lip-sync deepfakes. Table 1 presents a comparison between our dataset and recent datasets released over the past seven years.

We focus exclusively on “talking heads” in which one person is talking or gesturing in a typical video conferencing setup. We focus on this context because of the increasing prevalence of distinct harms that have emerged from both offline and real-time deepfake identity impersonations (including impostor hiring, fraud, and disinformation). This scenario presents unique challenges concerning identity verification and liveness detection, as compared to the more common analysis of text- or image-to-video content. The “talking heads” scenario shifts the question from only a ‘real v. fake’ or ‘identity’ question to both one of realism *and* identity. Our effort provides a dataset to simultaneously address both of these questions. Existing datasets in this domain do not reflect the breadth of ours, nor additionally, the current quality and diversity of deepfake generators, while bearing ethical, practical, and legal shortcomings (see Table 1). Specifically, existing datasets largely comprise low-quality, outdated deepfake generators, where the underlying data was scraped without participant consent. Moreover,

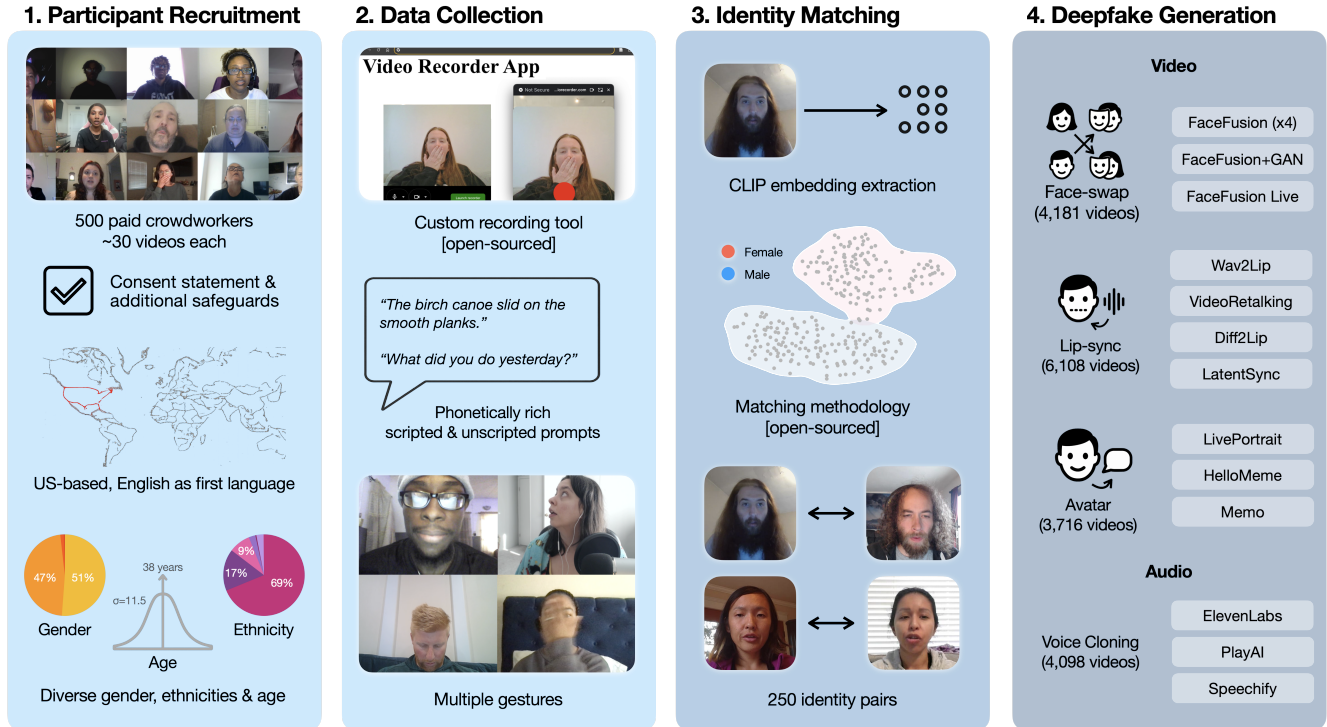


Figure 1. An overview of the *DeepSpeak* Dataset sourced from a diverse selection of consenting participants using a custom-built data collection methodology. The dataset also comprises deepfakes generated from 14 video and three audio deepfake methods using facial identity matching to improve the realism of the generated deepfakes.

these datasets do not include all types of deepfake generators and attack settings. A more comprehensive review of other datasets is included in Appendix B.

To remedy these shortcomings, our work makes the following contributions:

- **Documentation and Release of DeepSpeak.** We introduce a methodology for the large-scale collection of real video recordings, self-submitted by a diverse selection of consenting participants (Section 2), along with the procedures used to generate corresponding deepfake video and audio (Sections 4 and 5).
- **Data Collection Tool.** We provide a codebase for a web-based application designed to facilitate participant-led remote data collection. When used in conjunction with our collection survey, the collected data is phonetically rich and diverse in terms of speech content, video durations, gestures, and includes both scripted and unscripted segments.
- **Method for Identity Matching.** We devise a method for matching participants based on their visual features to create more convincing face-swap deepfakes, consistent with real-world deepfake attacks (Section 3).
- **Large-scale Benchmarking and Generalization Study.** We perform large-scale evaluations of state-of-the-art deep-

fake detectors across audio, visual and multimodal detectors and show that they fail to accurately distinguish between real and fake audio and video when trained on other datasets (Table 1)(Section 6). These evaluations highlight the importance of a large and diverse dataset containing deepfakes from the latest generative-AI tools.

2. Data Collection

The data collection was performed in four steps. Data collection for *DeepSpeak* was determined to qualify for exempt status by UC Berkeley Office for Protection of Human Subjects (OPHS).

Participant Recruitment. Participants were crowd-sourced through the Prolific research recruitment platform. Participants were asked to give their consent for including their recordings, without any other identifying information, in a public dataset. Details of the consent statement can be found in Appendix P. A total of 500 participants were selected from a stratified sample ensuring equal distribution of gender, and with all participants reported as being native English speakers and U.S. residents, with demographics as follows (some participants identified with more than one race/ethnicity):

Name	Release Year	Unique Identities	Original Footage	Consent	Faceswap	Lipsync	Avatar	Fake Audio	Conversational Webcam	Identity Matching	Deepfake Footage
FaceForensics	2018	NA	1,004	N	?	-	-	-	-	-	2,008
FaceForensics++	2019	NA	1,000	Partial	✓	-	-	-	-	-	4,000
DFDC	2020	3,426	23,654	Y	✓	-	-	-	Partial	✓	104,500
DFD	2019	28	363	Y	✓	-	-	-	✓	-	3,068
Celeb-DF	2020	59	590	N	✓	-	-	-	-	-	5,049
AVDeepfake1M	2024	2,068	286,721	N	-	✓	-	✓	-	-	860,039
FakeAVCeleb	2021	600	570	N	✓	✓	-	✓	-	-	25,000
Deepfake-Eval-2024	2025	NA	-	N	✓	-	-	✓	Partial	-	-
LAV-DF	2022	153	36,431	N	-	✓	-	✓	-	-	99,873
DF40	2024	NA	NA	N	✓	✓	✓	-	-	-	100,000+
NVFAIR	2025	161	NA	Y	-	✓	-	-	✓	-	650,000
Polyglotfake	2024	NA	766	N	-	✓	-	✓	-	-	14,472
Illusion	2025	NA	139,740	N	✓	-	-	✓	-	✓	1,232,246
DF-Platter	2023	454	764	N	✓	-	-	-	-	-	132,496
DeepSpeak (Ours)	2025	500	16,043	Y	✓	✓	✓	✓	✓	✓	14,005

Table 1. A comparison of forensic-themed public datasets. Although not the most informative metric, we report original and deepfake footage as number of videos for consistency with previous published datasets (NA: not available). “Fake Audio” refers to speech synthesized by AI-enabled voice cloning.

- **Age:** Range = 18-75 years, Mean = 38 years; standard deviation = 11.5 years
- **Gender:** 256 male, 235 female, 7 non-binary, 2 not provided
- **Race/Ethnicity:** 362 White/Caucasian, 87 Black/African American, 45 Asian, 14 American Indian/Alaska Native, 2 Native Hawaiian/Other Pacific Islander, 15 other, 1 prefer not to say.

Survey. The data collection survey was designed to capture both speech and visual actions. For speech, it included phonetically rich audio data spanning varied audio durations with both scripted vs. conversational-style responses. Each participant was instructed to record themselves responding to between 32 and 35 separate prompts. Participants were paid \$7 for their time. The first two prompts were used for voice-clone training data (see Section 4). The remaining prompts were divided into four categories: (1) 10 standardized scripted responses in which each participant read the same prompt; (2) 10 randomized scripted responses in which participants read a randomized prompt; (3) 10 unscripted responses in which participants responded to questions; and (4) between 5-8 actions in which participants performed simple actions. Scripted responses were generated using transcripts of the TIMIT dataset, a linguistics research dataset consisting of utterances from 462 real female and male American-English speakers. See Appendix Q for the full list of prompts and scripts used.

Data Collection tool. Both audio and video were recorded using a custom-built Google Chrome web application. The JavaScript and Python repository for this web application is available at <https://github.com/hfaridlab/>

deepspeak. Details of the encoding and data pre-processing associated with the tool can be found in the Appendix C.

Validation. Participants were given written and visual instructions to allow them to practice recording themselves and test their hardware. Participants were asked to adhere to a series of recording conditions intended to improve consistency within the overall dataset. We manually removed any invalid responses from the final dataset that did not meet these requirements. The details of this can be found in the Appendix C.

3. Identity Matching

During manual inspection of the collected data, we observed that, albeit diverse in age, gender, and ethnicity, our collected data contains many individuals with similar facial and vocal features. In order to exploit this feature of the dataset and create more compelling deepfakes, each identity in the dataset was paired with another, perceptually and acoustically similar one. The code for producing this visual and audio matching, as well as the resulting visual and audio pairs is open-sourced at <https://github.com/hfaridlab/deepspeak>.

Visual Matching. Each identity is first represented by the average CLIP embedding¹ [39] extracted from five random video frames (filtered for low-quality frames, see Section 5.2). Shown in Figure 2 is a t-SNE visualization of a subset of these embeddings. Comparing this representation against the self-reported demographic information reveals

¹<https://github.com/OpenAI/CLIP>

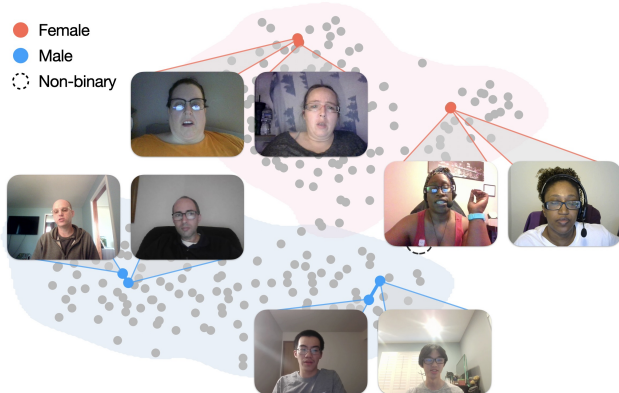


Figure 2. A t-SNE visualization of CLIP embeddings from real participant’s videos. The four highlighted pairs correspond to identities with maximal similarity as measured by the cosine distance between CLIP embeddings. Perceptually similar identities cluster in this t-SNE representation. The red/blue color coding corresponds to people who identify as female/male, which also clusters in this t-SNE representation.

that these CLIP embeddings cluster based on gender, ethnicity and facial similarity.

For each identity, a unique matched identity is assigned using the agglomerative clustering algorithm with cosine distance and cluster size constraint from the scikit-learn library². Additional examples of visual pairs are shown in Appendix I. This approach was adopted instead of a more traditional biometric matching like ArcFace [12] because we observed, during manual review, better qualitative matching for women and people of color. We also found that our CLIP-based matching outperforms ArcFace in terms of the Frechet inception distance (FID) between the matched face pairs by 4% (214 vs 222), and the LPIPS distance between the matched face pairs by 6% (0.61 vs 0.65). For both FID and LPIPS, a smaller value corresponds to higher perceptual similarity.

With this identity matching, averaged across all videos, DeepSpeak achieves an average FID of 238 and LPIPS of 0.46. Compared to baselines datasets, this is 36% better than DF40 (325 FID and 0.68 LPIPS), 33% better than FaceForensics (317 and 0.74), and 71% better than DFDC (408 and 0.70).

4. Audio Generation

Participants were first asked to record themselves reading 10 consecutive phonetically-rich sentences, sourced from List 1 of the standard Harvard Sentences [42], a collection of sentences representing best practice for standardized evaluation of speech processing and audio quality in controlled settings.

Participants were then asked to repeat the standard elicitation paragraph from the Speech Accent Archive, a phonetically comprehensive passage comprising a breadth of vowels and consonants [54]. These two scripted responses were used for the purpose of voice cloning, and had an average length of 30 seconds.

Using each participant’s cloned voice, a synthetic audio was created in their voice saying the same thing as in the original audio/video. For the unscripted responses, the original audio was transcribed using OpenAI’s Whisper, and for the scripted responses, we assumed that the participant correctly read the script. These text transcriptions were then provided to each voice cloning generators’ API to generate matching synthetic voices.

Voice clones were generated using three commercial cloning and Text-to-Speech (TTS) services: ElevenLabs, PlayAI and Speechify. The details of API end points used, alongside parameters, can be found in Appendix E.

5. Video Generation

We generated three types of video deepfakes: face swap, lip sync, and avatar, each of which is described next. The resulting dataset is randomly split into 80/20 training/testing splits with no overlap in facial or voice identities. A breakdown of the resulting dataset’s statistics, including the total file size (GB), file counts (N), and video length (hrs) are included in Appendix D.

5.1. Generation

Face-Swap. Face-swap deepfakes are created by replacing – eyebrow to chin and cheek to cheek – the original identity in a video with a new identity. We swapped faces of identity pairs identified through the visual matching (see Section 3). This ensured that the swapped identities were perceptually similar to begin with, which made for more compelling deepfakes. This resembles conventional practices of in-the-wild deepfake production, where actors are chosen based on their similarity to the target identity.

An overview of face-swap deepfake generation is shown in Figure 3, row one. To generate a face swap, the video of the original identity and a single frame of the matched identity are provided to the face-swap synthesis engine. The single frame is initially chosen to be the fifth frame in a randomly selected video of the matched identity. We found that if the eyes are closed in the matched face, the resulting face-swap deepfake suffered in quality. As such, we used MediaPipe [33] to extract facial features and ensured that the distance between the top and bottom eyelid landmarks was greater than a specified threshold. If this constraint failed, then the tenth frame was selected for consideration; this process was repeated, skipping five frames each time, until a suitable frame was found. We used seven face-swap methods, as detailed in Appendix D.

²<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>

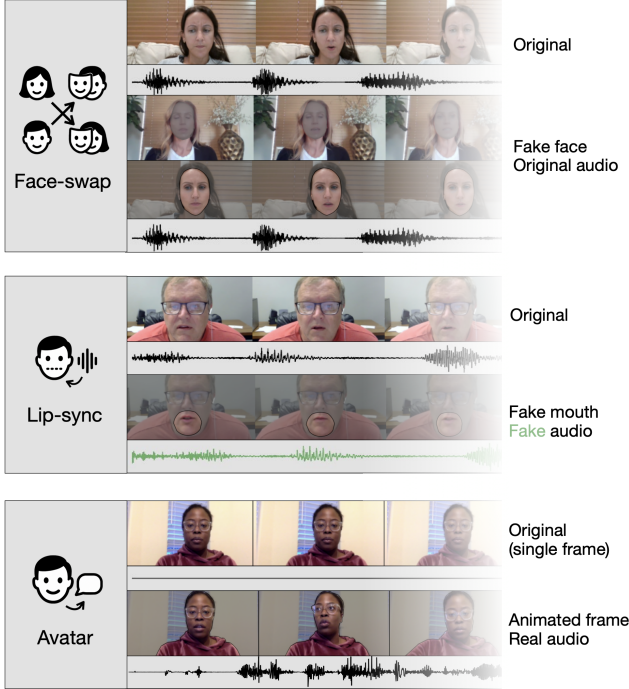


Figure 3. The *DeepSpeak* dataset consists of face-swap, lip-sync, and avatar deepfakes.

Lip-Sync. Whereas the face-swap deepfake replaces an entire face with a new identity, a lip-sync deepfake modifies the mouth region to be consistent with a different audio track. An overview of lip-sync deepfake generation is shown in row two of Figure 3. Given an original video and associated audio, we create two types of lip-sync deepfakes: (1) a lip-sync deepfake with an audio of the same identity extracted from a different video (i.e., the audio and video are now mismatched); and (2) a lip-sync deepfake with an AI-generated voice of the same identity (Section 4) with a transcript taken from a different video. Four methods of lip-sync deepfakes were employed, as further described in Appendix D.

Avatar. An overview of avatar deepfake generation is shown in Figure 3, row three. Avatar deepfakes animate the head and lip movements of a static image to match a target video or audio track. Unlike face-swap and lip-sync deepfakes, which modify an existing video, avatar deepfakes generate movement from a single static image. Avatar deepfakes were created using three methods further described in Appendix D. LivePortrait and HelloMeme take as input a single image of a person to animate with a video (and associated audio) that drives this animation. For these two generators, the avatar deepfakes contain only real audio from the original driving video. Memo takes as input a single image of a person to animate with only an audio that drives this animation. In this case, the audio can be either real or fake.

5.2. Validation

During manual inspection of the generated videos, we identified multiple types of failures, including deepfake engines (1) producing corrupted faces with consistently closed eyes or mouths, (2) generating malformed avatars with distorted facial or upper body structure, (3) failing to apply any changes and yielding back the original video, (4) modifying only parts of the video, (5) producing empty output consisting of with black frames, among others. To prevent failed deepfakes, we designed a suite of input and output detectors to filter undesired features. This filtering code is open-sourced at <https://github.com/hfaridlab/deepspeak>. The details of this filtering can be found in Appendix F.

6. Experiments

We conducted a series of baseline experiments on *DeepSpeak* for the tasks of audio and video deepfake detection. The code for these experiments, including data pre-processing, is open-sourced at <https://github.com/hfaridlab/deepspeak>. The experiments were conducted on NVIDIA A100 GPUs over the course of approximately four weeks (see Appendix O for details pertaining compute resources).

6.1. Video Deepfake Detection

Baselines. Both classic- and deep-learning methods for deepfake video detection can be categorized by the scrutinized signal deemed to discriminate the real from the fake, with most performing (1) spatial-domain analysis, (2) frequency-domain analysis, or (3) cross-modal temporal coherence analysis. To capture the breadth of the existing approaches, we evaluate state-of-the-art methods representing these distinct lines of work. The first evaluated architecture is a frequency-based method FreqNet [7]. The second, spatial-domain, architecture is GenConViT [55] (with ED and VAE variants). The third, multi-modal, architecture is LipFD [30] designed to detect misalignments between the visual and vocal stream of lip-sync deepfakes.

For each of these four architectures, we evaluated three model variants: (1) the pretrained model released alongside the respective publication (trained on a different, non-*DeepSpeak* dataset), (2) the model trained from scratch on *DeepSpeak*, and (3) the model, starting with the pre-trained weights, fine-tuned on *DeepSpeak*. A total of 12 models were evaluated.

Experimental Setup. To perform inference, training, and fine-tuning of the included architectures, we used the official code repositories released alongside the respective publications. To make the results comparable despite the differing number of parameters of these architectures, we used default hyperparameters when possible, with a simple search over learning rates (see Appendix L for details).

Each model is evaluated against the testing split of its architecture’s original dataset and DeepSpeak. The original dataset refers to the dataset used for the pretrained model in the respective publication: for FreqNet, it is a custom GAN-generated dataset compiled by its authors [7]; for GenConViT ED and VAE, it is Celeb-DF 2 [28]; and for LipFD, it is AVLips [30]. The accuracy on the real and fake class is reported separately, along with the overall F1 score.

Results. Shown in Table 2 are the results of the pretrained models and models trained from scratch on DeepSpeak. All four evaluated architectures follow the same pattern: they perform reasonably well on the testing splits of their original training datasets but fail to generalize to DeepSpeak. The same trend holds when models are trained on DeepSpeak and evaluated on the original dataset. Notably, even on the original testing sets, class bias was evident—for example, GenConViT attained an accuracy of 98.2% on fake but only 56.7% on real, while LipFD showed the opposite pattern, scoring 97.9% on real versus 69.1% on fake.

Also shown in last six columns of Table 2 are the results of the fine-tuned models (labeled Original + DeepSpeak). While some models, such as GenConViT ED and VAE, achieved performance on DeepSpeak comparable to training from scratch (F1 score above 0.9), this came at the cost of a sharp drop in performance on the original testing set, where F1 scores fell below 0.2. LipFD was able to fine-tune on DeepSpeak while maintaining comparable performance on the original testing set (both with F1 scores around 0.7), though it should be noted that the model exhibits a strong bias toward the fake class.

6.2. Audio Deepfake Detection

Baselines. We evaluated the performance of two model architecture types on the DeepSpeak dataset, consistent with recent literature: (i) a foundation model, and (ii) a raw waveform model. Foundation models use a pretrained model to extract embeddings from the input waveform, which are then passed to a classifier. Three state-of-the-art models were selected: TitaNet [3, 25], Wav2Vec-XLSR [2, 37], and LAION-CLAP [37, 56]. For each embedding type, both linear and non-linear classifiers were tested. Raw waveform models operate directly on the audio waveform. Three leading models were chosen: AASIST [23], RawNet2 [23, 48], and RawGAT-ST [23, 47].

For both architectures, we evaluated two versions of each model: (1) a pretrained model trained on a dataset other than DeepSpeak, and (2) a model trained from scratch on DeepSpeak. In the case of foundation models, the foundation model used to extract embeddings remained pretrained, while the downstream classifiers were trained from scratch. In total, 18 models were evaluated. A summary is provided in Table 6.2.

Experimental Setup. Pretrained raw waveform model weights were sourced directly from the AASIST implementation of AASIST, RawNet2, and RawGAT-ST³. Default configuration were used for each model, as detailed in the Appendix L. For retraining these models from scratch on DeepSpeak, the same configurations and architectures were maintained, with DeepSpeak training data replacing ASVSpooof. For foundation models, classifiers (both logistic regression and random forest) were trained using balanced datasets with embeddings extracted from the training sets of either ASVSpooof (for Wav2Vec-XLSR and LAION-CLAP) or TIMIT-ElevenLabs (for TitaNet). Embeddings from the DeepSpeak test dataset split were used for evaluation. Both linear (logistic regression) and non-linear (random forest) classifiers were tested for each embedding type. No cross-validation or hyperparameter tuning was performed for either the pretrained or from-scratch models.

Each model is evaluated against the testing split of its architecture’s original dataset and DeepSpeak. The original dataset corresponds to the one used for pretraining in the respective publications. For AASIST, RawNet2, and RawGAT-ST, this dataset is ASVSpooof (as implemented in [23]), and for TitaNet-based embeddings approaches, this dataset is TIMIT-ElevenLabs [3]. Since prior literature on detection using Wav2Vec-XLSR and LAION-CLAP largely focusses on training-free methods [37], we trained our own benchmarks on ASVSpooof for consistency and because it serves as one of the most comprehensive and widely used benchmarking datasets. Performance metrics are reported for both the original dataset’s test set and the DeepSpeak test set. As shown in Table 6.2, accuracies for both real and fake classes are presented separately, along with overall accuracy to account for class imbalance (since fake audio only occurs in lip-sync deepfakes, representing a subset of the full dataset), and the error rate (EER).

Results. When trained and tested on DeepSpeak, raw waveform models perform well, with AASIST achieving 98.8% accuracy - only 0.7 percentage points lower than its original ASVSpooof benchmark. Embedding-based models also show strong, though comparatively lower performance, with the best performing models being those trained on LAION-CLAP embeddings (see Table 6.2).

Pretrained models, however, do not generalize well to DeepSpeak data. AASIST remains the top-performing pretrained model, albeit with substantially lower performance when evaluated out-of-the-box on DeepSpeak data, dropping to an accuracy of 60.1% and 60.2% for real and fake. Pretrained embedding-based models also show substantially lower performance when evaluated on DeepSpeak data, alongside notable class imbalances (see Table 6.2).

These results suggest that feature representations learned

³<https://github.com/clovaai/aasist>

Method	Original						DeepSpeak						Original + DeepSpeak					
	Original			DeepSpeak			Original			DeepSpeak			Original			DeepSpeak		
	Real	Fake	F1	Real	Fake	F1	Real	Fake	F1	Real	Fake	F1	Real	Fake	F1	Real	Fake	F1
FN	97.1	88.3	0.9	65.3	15.4	0.2	34.4	26.6	0.6	77.3	69.9	73.6	50.5	14.1	0.3	74.2	66.1	0.7
GC-ED	57.3	98.2	0.7	88.5	39.1	0.7	2.8	100	0.1	90.5	90.7	0.9	7.9	100	0.2	91.7	78.2	0.9
GC-VAE	56.7	98.2	0.7	88.5	39.2	0.7	4.5	100	0.1	91.1	96.4	0.9	9.0	99.7	0.2	93.0	89.6	0.9
LFD	97.9	69.1	0.8	98.8	3.5	0.1	7.30	88.7	0.7	71.8	77.1	0.8	2.8	97.8	0.7	28.2	96.6	0.7

Table 2. **Video deepfake detection accuracies (%)** of four state-of-the-art architectures: FreqNet (FN), GenConViT ED (GC-ED), GenConViT VAE (GC-VAE), and LipFD (LFD). The heading in the first row corresponds to the dataset on which each model was trained, and the heading in the second row corresponds to the dataset on which each model is evaluated against.

Model	Clf	Original						DeepSpeak					
		Original			DeepSpeak			Original			DeepSpeak		
		Real	Fake	F1	Real	Fake	F1	Real	Fake	F1	Real	Fake	F1
Titanet (FM)	LR	99.4	100.0	1.00	10.0	97.4	0.18	61.6	98.8	0.75	91.3	89.1	0.95
	RF	99.8	100.0	1.00	54.2	64.3	0.68	74.8	83.1	0.71	96.3	79.3	0.97
Wav2Vec2-xlsrc (FM)	LR	79.7	82.7	0.48	1.3	97.0	0.03	7.6	83.8	0.06	76.8	65.6	0.84
	RF	98.1	88.5	0.66	19.3	95.4	0.32	93.6	36.9	0.25	97.4	78.0	0.97
LAION-CLAP (FM)	LR	92.9	92.0	0.71	33.8	76.6	0.49	90.3	53.3	0.30	93.7	91.9	0.96
	RF	93.1	90.5	0.67	65.3	68.3	0.77	93.9	56.7	0.33	95.8	89.6	0.97
AASIST (RW)	-	99.5	99.5	0.98	60.1	60.2	0.72	73.1	73.1	0.36	98.8	98.8	0.99
RawNet2 (RW)	-	99.0	99.0	0.95	54.4	54.4	0.67	69.4	69.3	0.32	94.1	94.3	0.96
RawGAT-ST (RW)	-	99.1	99.1	0.96	57.7	57.7	0.70	75.0	75.0	0.38	96.8	96.8	0.98

Table 3. **Audio deepfake detection accuracies (%)** of nine state-of-the-art models using two separate architectures (FM = Foundation Model, RW = Raw Waveform). For each training/testing combination, we report the real class accuracy, fake class accuracy, and F1 score. The Clf column indicates the type of classifier used for embedding-based models, either logistic regression (LR) or random forest (RF).

directly from raw waveform inputs may be more resilient to domain shift in DeepSpeak data than those extracted from foundation embeddings-based models. In all cases, pre-trained models are insufficient for accurately distinguish real from fake audio.

This pattern for audio detection models is similar to video detection models: (1) these models struggle with out-of-domain data; but (2) these models can improve with appropriate training.

6.3. Combined Audio-Visual Detection

While most deepfake detection methods treat audio and video independently, several joint approaches have also been proposed. Although multimodal detection is not the primary focus of this work, we confirm that the limited generalization observed in unimodal detection methods extends to a leading combined method exploiting audio-video mismatch [6]. This technique classifies a video as real/fake based on a normalized Levenshtein distance between an audio transcription and a video transcription (based on an automatic lip reading). At a threshold of 0.55 on this distance (where are distance of 0.0 corresponds to a perfect match between the audio and video, and a distance of 1.0 corresponds to a

maximal mismatch), the accuracy across 2963 real and 2953 AI-generated videos is 65.5% and 67.9%. Accuracy across different types of AI-generated videos (face-swap, lip-sync, and avatar) ranged from a low of 43.4% to a high of 99.0%.

7. Closing Thoughts

Discussion. Year after year, we see a dramatic rise in the number of deepfake generators and the quality of the fake audio and video. Given the pace at which deepfake technology is progressing, it is critical that evaluation datasets keep up with the latest technologies. This is made apparent by our evaluation of recent state-of-the-art deepfake detectors that struggle to generalize to the latest deepfake generators. To this end, our DeepSpeak dataset is partitioned into two parts (v1 and v2), containing a snapshot of the state of the art in deepfake generation in 2024 and 2025, respectively. We plan to release one to two new datasets each year to keep pace with these new threats.

Limitations. *DeepSpeak* captures the state of the art in deepfake generation at the time of publication, making it well-suited for developing and evaluating detection methods for current and emerging deepfake engines. However, as

generative AI evolves rapidly, it is essential to recognize the dataset’s limitations and the potential for future expansion.

Due to the lack of high-quality open-source deepfake engines for non-English languages, DeepSpeak currently includes participants speaking English. As high-quality multilingual engines become available, we will need to expand DeepSpeak to include additional languages.

Currently, open-source deepfake engines operate on the video level, which is reflected in DeepSpeak—every video in the dataset is either entirely real or entirely fake. Once targeted manipulation (i.e., changing only some words in the video) improves, we will include them in future versions.

Lastly, to date, all of the DeepSpeak video generators are based on open-source models and not on commercially available models. As we have done with commercial audio generators, we will seek to establish relationships with commercial video generators to allow for large-scale video generation of commercial offerings.

Ethical Considerations: Too many other datasets in media forensics and computer vision have adopted a “scrape and distribute, ask questions later” approach. We take issue with this both from the perspective of participant consent and intellectual property.

While we don’t object to the development of deepfake generators, we will not knowingly license *DeepSpeak* for this purpose. Our rationale here is that the harms that are coming from deepfakes are not insignificant and we simply don’t want to be contributing to a plethora of online harms.

Conclusion. Our motivation for creating this dataset is to support the media-forensics research community and the development and refinement of techniques to detect deepfake audio, image, and video. The world of generative AI and media forensics is fast moving. It is, therefore, important that shared datasets be regularly updated to keep up with the latest trends. To this end, we expect to release updates to this dataset once to twice a year. To help serve the community better, we welcome feedback, comments, requests for future releases of this dataset at <https://github.com/hfaridlab/deepspeak>.

Acknowledgments

We are grateful to David Chan for his many insightful comments and suggestions that significantly improved the quality of this paper. This work was supported by Google/YouTube and the University of California Noyce Initiative. We are grateful to ElevenLabs (<https://elevenlabs.io>) and PlayAI (<https://play.ai/>) for granting us API access for voice generation.

References

- [1] Triantafyllos Afouras, Joon Son Chung, Andrew Senior, Oriol Vinyals, and Andrew Senior. Deep audio-visual speech

- recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8717–8727, 2018. 17
- [2] Arun Babu, Changan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, and Michael Auli. XLS-R: Self-supervised cross-lingual speech representation learning at scale. In *Interspeech*, pages 2278–2282, 2022. 6
- [3] Sarah Barrington, Romit Barua, Gautham Koorma, and Hany Farid. Single and multi-speaker cloned voice detection: From perceptual to learned features. In *IEEE International Workshop on Information Forensics and Security*, pages 1–6. IEEE, 2023. 6
- [4] Sarah Barrington, Emily A Cooper, and Hany Farid. People are poorly equipped to detect AI-powered voice clones. *Scientific Reports*, 15(1):11004, 2025. 1, 30
- [5] Jon Bateman. *Deepfakes and synthetic media in the financial system: Assessing threat scenarios*. Carnegie Endowment for International Peace., 2022. 1
- [6] Matyas Bohacek and Hany Farid. Lost in translation: Lip-sync deepfake detection from audio-video mismatch. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4315–4323, 2024. 7
- [7] Runyuan Cai, Yue Ding, and Hongtao Lu. FreqNet: A frequency-domain image super-resolution network with discrete cosine transform. 2021. 5, 6
- [8] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. Simswap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2003–2011, 2020. 16
- [9] Kun Cheng, Xiaodong Cun, Yong Zhang, Menghan Xia, Fei Yin, Mingrui Zhu, Xuan Wang, Jue Wang, and Nannan Wang. VideoRetalking: Audio-based lip synchronization for talking head video editing in the wild. In *SIGGRAPH Asia*, pages 1–9, 2022. 17
- [10] Bobby Chesney and Danielle Citron. Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, 107:1753, 2019. 1
- [11] Joon Son Chung, Arsha Nagrani, and Andrew Senior. VoxCeleb2: Deep speaker recognition. arXiv:1806.05622, 2018. 17
- [12] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019. 4
- [13] Michelle L Ding and Harini Suresh. The malicious technical ecosystem: Exposing limitations in technical governance of ai-generated non-consensual intimate images of adults. arXiv:2504.17663, 2025. 1
- [14] Brian Dolhansky, Joanna Bitton, Ben Pfau, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer. The deepfake detection challenge (DFDC) dataset. arXiv:2006.07397, 2020. 14
- [15] Nick Dufour and Andrew Gully. Contributing data to deepfake detection research. Google AI Blog, 2019.
- [16] Emilio Ferrara. Charting the landscape of nefarious uses of generative artificial intelligence for online election interference. arXiv:2406.01862, 2024. 1

- [17] J. H. Frank and L. Schönherr. WaveFake: A Data Set to Facilitate Audio DeepFake Detection. In *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, pages 1–18, 2021. 14
- [18] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren. DARPA TIMIT acoustic phonetic continuous speech corpus, 1993. 15
- [19] Matthew Groh, Ziv Epstein, Chaz Firestone, and Rosalind Picard. Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, 119(1):e2110013119, 2022. 1
- [20] Jianzhu Guo, Dingyun Zhang, Xiaoqiang Liu, Zhizhou Zhong, Yuan Zhang, Pengfei Wan, and Di Zhang. LivePortrait: Efficient portrait animation with stitching and retargeting control. 2024. 17
- [21] Keith Ito and Linda Johnson. The lj speech dataset. <https://keithito.com/LJ-Speech-Dataset/>, 2017. 14
- [22] Ye Jia, Yu Zhang, Ron J. Weiss, Quan Wang, Jonathan Shen, Fei Ren, Zhifeng Chen, Patrick Nguyen, Ruoming Pang, Ignacio Lopez Moreno, and Yonghui Wu. Transfer learning from speaker verification to multispeaker text-to-speech synthesis. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 4485–4495, Red Hook, NY, USA, 2018. Curran Associates Inc. 14
- [23] Jee-weon Jung, Hee-Soo Heo, Hemlata Tak, Hye-jin Shim, Joon Son Chung, Bong-Jin Lee, Ha-Jin Yu, and Nicholas Evans. AASIST: Audio anti-spoofing using integrated spectro-temporal graph attention networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6367–6371, 2022. 6, 31
- [24] Hasam Khalid, Shahroz Tariq, Minha Kim, and Simon S Woo. FakeAVCeleb: A novel audio-video multimodal deepfake dataset. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. 14
- [25] Nithin Rao Koluguri, Taejin Park, and Boris Ginsburg. TitaNet: Neural model for speaker representation with 1D depth-wise separable convolutions and global context. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8102–8106. IEEE, 2022. 6
- [26] Prajwal KR, Rudrabha Mukhopadhyay, Jerin Philip, Abhishek Jha, Vinay Nambodiri, and CV Jawahar. Towards automatic face-to-face translation. In *27th ACM International Conference on Multimedia*, pages 1428–1436, 2019. 17
- [27] Chunyu Li, Chao Zhang, Weikai Xu, Jinghui Xie, Weiguo Feng, Bingyue Peng, and Weiwei Xing. LatentSync: Audio conditioned latent diffusion models for lip sync. 2024. 17
- [28] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. Celeb-DF: A large-scale challenging dataset for deepfake forensics. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3207–3216, 2020. 6, 14
- [29] Mingcong Liu, Qiang Li, Zekui Qin, Guoxin Zhang, Pengfei Wan, and Wen Zheng. BlendGAN: Implicitly GAN blending for arbitrary stylized face generation. *Advances in Neural Information Processing Systems*, 34:29710–29722, 2021. 17
- [30] Weifeng Liu, Tianyi She, Jiawei Liu, Boheng Li, Dongyu Yao, and Run Wang. Lips are lying: Spotting the temporal inconsistency between audio and visual in lip-syncing deepfakes. *Advances in Neural Information Processing Systems*, 37:91131–91155, 2024. 5, 6
- [31] Xuechen Liu, Xin Wang, Md Sahidullah, Jose Patino, Héctor Delgado, Tomi Kinnunen, Massimiliano Todisco, Junichi Yamagishi, Nicholas Evans, Andreas Nautsch, and Kong Aik Lee. Asvspoof 2021: Towards spoofed and deepfake speech detection in the wild. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 31:2507–2522, 2023. 14
- [32] Steven R Livingstone and Frank A Russo. The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in north american english. *PloS one*, 13(5):e0196391, 2018. 17
- [33] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. MediaPipe: A framework for building perception pipelines. arXiv:1906.08172, 2019. 4, 18
- [34] Soumik Mukhopadhyay, Saksham Suri, Ravi Teja Gadde, and Abhinav Shrivastava. Diff2Lip: Audio conditioned diffusion models for lip-synchronization. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5292–5302, 2024. 17
- [35] Arsha Nagrani, Joon Son Chung, and Andrew Zisserman. VoxCeleb: A large-scale speaker identification dataset. arXiv:1706.08612, 2017. 17
- [36] Sophie J. Nightingale and Hany Farid. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, 119(8):e2120481119, 2022. 1
- [37] Alessandro Pianese, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Training-free deepfake voice recognition by leveraging large-scale pre-trained models. In *ACM Workshop on Information Hiding and Multimedia Security*, page 289–294, New York, NY, USA, 2024. 6
- [38] KR Prajwal, Rudrabha Mukhopadhyay, Vinay P Nambodiri, and CV Jawahar. A lip sync expert is all you need for speech to lip generation in the wild. In *28th ACM International Conference on Multimedia*, pages 484–492, 2020. 17
- [39] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021. 3
- [40] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR, 2023. 17
- [41] Xingyu Ren, Alexandros Lattas, Baris Gecer, Jiankang Deng, Chao Ma, and Xiaokang Yang. Facial geometric detail recovery via implicit representation. In *IEEE 17th International Conference on Automatic Face and Gesture Recognition*, 2023. 15, 16
- [42] Ernst H Rothaus. Ieee recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3):225–246, 1969. 4

- [43] Henry Ruhs. FaceFusion. <https://github.com/facefusion/facefusion>, 2024. 15
- [44] Ryosuke Sonobe, Shinnosuke Takamichi, and Hiroshi Saruwatari. JSUT corpus: Free large-scale Japanese speech corpus for end-to-end speech synthesis, 2017. arXiv preprint. 14
- [45] Sam Stockwell, Megan Hughes, Phil Swatton, Albert Zhang, Jonathan Hall KC, and Kieran. AI-enabled influence operations: Safeguarding future elections. Technical report, Centre for Emerging Technology and Security (CETaS), The Alan Turing Institute, 2024. 1
- [46] Kim Sung-Bin, Lee Chae-Yeon, Gihun Son, Oh Hyun-Bin, Janghoon Ju, Suekyeong Nam, and Tae-Hyun Oh. MultiTalk: Enhancing 3D talking head generation across languages with multilingual video dataset. arXiv:2406.14272, 2024. 17
- [47] Hemlata Tak, Jee-Weon Jung, Jose Patino, Madhu Kamble, Massimiliano Todisco, and Nicholas Evans. End-to-end spectro-temporal graph attention networks for speaker verification anti-spoofing and speech deepfake detection. arXiv:2107.12710, 2021. 6
- [48] Hemlata Tak, Jose Patino, Massimiliano Todisco, Andreas Nautsch, Nicholas Evans, and Anthony Larcher. End-to-end anti-spoofing with RawNet2. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6369–6373, 2021. 6
- [49] Cristian Vaccari and Andrew Chadwick. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, 6(1):2056305120903408, 2020. 1
- [50] Marco Viola and Cristina Voto. Designed to abuse? deepfakes and the non-consensual diffusion of intimate images. *Synthese*, 201(1):30, 2023. 1
- [51] Haofan Wang. INSwapper: Face swapping model based on insightface. <https://github.com/haofanwang/inswapper>, 2023. 16
- [52] Kaisiyuan Wang, Qianyi Wu, Linsen Song, Zhuoqian Yang, Wayne Wu, Chen Qian, Ran He, Yu Qiao, and Chen Change Loy. Mead: A large-scale audio-visual dataset for emotional talking-face generation. In *European Conference on Computer Vision*, pages 700–717. Springer, 2020. 17
- [53] Zhouxia Wang, Jiawei Zhang, Tianshui Chen, Wenping Wang, and Ping Luo. RestoreFormer++: Towards real-world blind face restoration from undegraded key-value pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 16
- [54] Steven Weinberger. Speech accent archive, 2015. Retrieved from the Speech Accent Archive. 4
- [55] Deressa Wodajo, Solomon Atnafu, and Zahid Akhtar. Deepfake video detection using generative convolutional vision transformer. 2023. 5
- [56] Yusong Wu, Ke Chen, Tianyu Zhang, Yuchen Hui, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1–5. IEEE, 2023. 6
- [57] Liangbin Xie, Xintao Wang, Honglun Zhang, Chao Dong, and Ying Shan. VFHQ: A high-quality dataset and benchmark for video face super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 657–666, 2022. 17
- [58] Zhiyuan Yan, Taiping Yao, Shen Chen, Yandan Zhao, Xinghe Fu, Junwei Zhu, Donghao Luo, Chengjie Wang, Shouhong Ding, Yunsheng Wu, et al. DF40: Toward next-generation deepfake detection. arXiv:2406.13495, 2024. 14
- [59] Shuang Yang, Yuanhang Zhang, Dalu Feng, Mingmin Yang, Chenhao Wang, Jingyun Xiao, Keyu Long, Shiguang Shan, and Xilin Chen. LRW-1000: A naturally-distributed large-scale benchmark for lip reading in the wild. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8. IEEE, 2019. 17
- [60] Shengkai Zhang, Nianhong Jiao, Tian Li, Chaojie Yang, Chenhui Xue, Boya Niu, and Jun Gao. HelloMeme: Integrating spatial knitting attentions to embed high-level and fidelity-rich conditions in diffusion models. 2024. 17
- [61] Zhimeng Zhang, Lincheng Li, Yu Ding, and Changjie Fan. Flow-guided one-shot talking face generation with a high-resolution audio-visual dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3661–3670, 2021. 17
- [62] Longtao Zheng, Yifan Zhang, Hanzhong Guo, Jiachun Pan, Zhenxiong Tan, Jiahao Lu, Chuanxin Tang, Bo An, and Shuicheng Yan. MEMO: Memory-guided diffusion for expressive talking video generation. 2024. 17
- [63] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *Advances in Neural Information Processing Systems*, 35:30599–30611, 2022. 16
- [64] Shangchen Zhou, Kelvin C.K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *NeurIPS*, 2022. 16, 17
- [65] Hao Zhu, Wayne Wu, Wentao Zhu, Liming Jiang, Siwei Tang, Li Zhang, Ziwei Liu, and Chen Change Loy. CelebV-HQ: A large-scale video facial attributes dataset. In *European Conference on Computer Vision*, pages 650–667. Springer, 2022. 17

Appendix

Table of Contents

A Release and Usage Information	13
B Related Work	14
B.1. Video	14
B.2. Audio	14
C Data Collection Details	14
C.1. Real Participants	14
C.2. Survey	15
C.3. Data Collection Tool	15
C.4. Real Video Validation	15
D Video Generation Methods	15
D.1. Face-Swap	15
D.2. Lip-Sync	17
D.3. Avatar	17
E Audio Generation Methods	17
F. Video Validation	18
G Video Statistics	18
H Deepfake Video Frame Examples	19
I. Visual Identity Pairing Examples	26
J. Audio Identity Matching	30
J.1. Audio Identity Matching Approach	30
J.2. Audio Identity Pairs	30
K Experimental Setup: Data Preprocessing	30
K.1. Video Deepfake Detection	30
K.2. Audio Deepfake Detection	31
L Experimental Setup: Hyperparameters	31
L.1. Video Deepfake Detection	31
L.2. Audio Deepfake Detection	31
L.3. Audio Raw Waveform Pretrained Model Configurations	33
L.4. Audio Raw Waveform Model Configurations for Training on DeepSpeak	35
MSafeguards against Misuse	35
N Licensing	35
N.1. DeepSpeak Dataset	35
N.2. Video Deepfake Detection Experiments	36

N.3 Audio Deepfake Detection Experiments	36
Q Compute Resources	36
O.1 DeepSpeak Dataset	36
O.2 Video Deepfake Detection Experiments	36
O.3 Audio Deepfake Detection Experiments	37
P. Survey Materials	38
Q Prompts	41
Q.1 Voice Cloning Prompts	41
Q.2 Standardized Scripted Prompts	41
Q.3 Unscripted Prompts	41
Q.4 Video Action Prompts	42
Q.5 Example frames from action prompts	43
Q.6 Randomized Scripted Prompts	44
R Video Identity Matching	52

A. Release and Usage Information

We released the DeepSpeak dataset in three separate batches: versions 1.0, 1.1 and 2.0. Additional real data was collected between versions 1.x and 2.0 (220 and 280 identities respectively). The training and testing splits of each version are detailed in Table 4. Furthermore, different generation engines were used between versions 1.x and 2.0 to reflect the current state-of-the-art methods at the time of release. The details of which engines were used in each version are shown in Table 5. Version 1.1 corrects for minor errors from version 1.0. As such, we recommend combining versions 1.1 and 2.0 when creating the complete dataset.

- Version 1.0: <https://huggingface.co/datasets/faridlab/deepspeak.v1>
- Version 1.1: <https://huggingface.co/datasets/faridlab/deepspeak.v1.1>
- Version 2.0: <https://huggingface.co/datasets/faridlab/deepspeak.v2>

Ver	Total			Train		Test	
	size (GB)	size (N)	size (hrs)	real (N [hrs])	fake (N [hrs])	real (N [hrs])	fake (N [hrs])
1.0	40	13,025	44.3	4,902 [13.9]	5,300 [21.0]	1,324 [3.7]	1,499 [5.8]
1.1	46	13,463	48.0	5,251 [16.8]	5,299 [21.0]	1,416 [4.4]	1,497 [5.8]
2.0	124	16,585	52.7	7,513 [23.6]	5,793 [18.6]	1,863 [5.8]	1,416 [4.6]

Table 4. A breakdown of the total size (gigabytes (GB), number of files (N), and length in hours (hrs)) of each version of the DeepSpeak dataset.

Ver	Audio	Face-swap	Lip Sync	Avatar
1.x	ElevenLabs	FaceFusion FaceFusion + <i>GAN</i> FaceFusion Live	Wav2Lip VideoRetalking	—
2.0	ElevenLabs PlayAI Speechify	INSwapper INSwapper + <i>CodeFormer</i> SimSwap SimSwap + <i>RestoreFormer</i>	Diff2Lip LatentSync	LivePortrait HelloMeme Memo

Table 5. An overview of deepfake generation engines used in each release version of the dataset.

B. Related Work

B.1. Video

Table 1 presents an overview of existing video deepfake datasets. None of these datasets contain all types of deepfakes (face-swap, lip-sync, avatar, and audio); the majority did not obtain consent from the individuals featured. Notable datasets are further detailed below.

DFDC. Released in 2020 as part of the DeepFake Detection Challenge, DFDC [14] contains 128,154 face-swap deepfakes of 3,426 paid, consenting actors. While the dataset brought significant attention to the problem of deepfake detection, it also sparked controversy due to failure cases (e.g., videos where the face-swap failed but were still labeled as deepfakes) and inconsistent annotations. Today, only a small subset of the labels is publicly available.

Celeb-DF. While most prior datasets were scripted and studio-recorded, Celeb-DF [28] aimed to mimic the in-the-wild nature of deepfake detection. It includes 590 real and 5,639 deepfake videos of 59 individuals who did not provide consent for such use. These were celebrities, mostly taken from YouTube videos. Like DFDC, the deepfakes were generated using face-swap models, but unlike DFDC, Celeb-DF’s annotations are consistent and fully available.

DF40. DF40 [58] contains over 100,000 deepfake videos spanning various types (lip-sync, face-swap, and avatar), featuring individuals who did not provide consent for inclusion. Identity and real video statistics are not reported. In creating the dataset, the authors collected some new data and repurposed content from Celeb-DF, FFHQ, and other datasets.

B.2. Audio

Most existing audio deepfake and spoofing datasets are single-modal, focusing exclusively on audio. The key datasets in this area are outlined below.

ASVspoof. ASVspoof [31] is considered one of the most popular audio spoofing datasets, and is commonly used for training and evaluating deepfake detection models. There have been multiple releases of this dataset (including 2019 and 2021), released alongside the ASVspoof challenges for each corresponding year and updated in accordance with new tools and generation methods. The dataset contains three categories of data: Physical Access (pertaining to audio undergoing physical attack methods such as replay attacks), Logical Access (pertaining to audio created by Text-To-Speech and voice conversion systems), and Deepfake Audio (as with Logical Access, but with generalized compression and codec variation).

While ASVspoof is considered a popular benchmarking dataset, it poses three main issues that we sought to address through releasing DeepSpeak: (1) the TTS and VC systems do not leverage state-of-the-art commercial platforms that are popular with real-world adversaries; (2) there are few real speakers used in the training and validation sets (for example, 20 speakers in ASVspoof 2021); and (3) the audio is not paired with video.

WaveFake. WaveFake [17] contains approximately 196 hours of both real and fake audio. The dataset is primarily based on the LJSPEECH dataset [21] (a public English speech corpus), alongside the JSUT dataset [44] (a Japanese speech corpus). Both of these datasets include audio clips recorded by a single female speaker. As such, WaveFake poses two additional issues that we sought to address through releasing DeepSpeak: (1) only comprising two speaker identities; (2) providing largely scripted audios rather than conversational; (3) generation methods that are no longer considered state-of-the-art (including MelGAN and HiFi-Gan methods).

FakeAVCeleb. FakeAVCeleb [24] is a multi-modal deepfake dataset. By way of fake audios, the dataset only comprises one generation method using a real-time voice cloning tool (SV2TTS [22], released in 2019). By contrast, DeepSpeak encompasses three more recently released state-of-the-art commercial voice clone and accompanying TTS methods.

C. Data Collection Details

C.1. Real Participants

Participants were asked to give their consent for including their recordings, without any other identifying information, in a public dataset. The precise consent language was: “This dataset will be used for research purposes for detecting deepfakes.

Please note that your recordings will be made public in a dataset, but no other identifying information will be shared outside of our research group. Please select the option below to consent to participate in this study.” The complete introductory page, including the consent information presented to participants, is available in Appendix P.

C.2. Survey

For scripted responses, participants were asked to record themselves repeating a short script while looking into the camera. Scripted responses were obtained using transcripts of the TIMIT dataset [18]. The TIMIT dataset consists of 462 real female and male American-English speakers, uttering a total of 1,718 short-to-medium length phonetically-rich sentences. Sentences of length less than the mean of 50 characters were removed. Ten sentences were then selected at random for the standardized scripts read by all participants. The remaining 728 sentences comprised the randomized scripts, for which each participant read a random sample of 10.

By way of unscripted prompts, participants responded to four open-ended unscripted questions and were asked to aim for a response that was close to 30 seconds in length. These were followed by quick-fire questions in which they repeated the question and provided a short response. Following the scripted and unscripted responses, participants were asked to perform simple actions using head and hand gestures.

By way of action prompts, participants were asked to perform seven simple visual actions: (1) wave their hand in front of their face while counting 1,2,3; (2) look down and right, straight down, and down and left, each time holding and counting 1,2,3; (3) look up and right, straight up, and up and left, each time holding and counting 1,2,3; (4) lean towards the camera without counting; (5) pretending to yawn; (6) pretending to laugh for between 2 to 5 seconds; (7) clapping loudly three times, pausing for about 5 seconds between claps.

The full list of all prompts, including scripted, unscripted and action-based, can be found in Appendix Q.

C.3. Data Collection Tool

Both audio and video were recorded using a custom-built Google Chrome web application. Recordings were captured as .webm files with a bitrate of 8 megabits per second, using the Google VP9 codec for video compression. The target resolution was set at 1280×720 pixels, but users with limited bandwidth were able to record at lower resolution of 640×480 pixels. The JavaScript and Python repository for this web application is available at <https://github.com/hfaridlab/deepspeak>. A screenshot of the recording web application is shown in 4.

The audio/video recordings captured by the tool were then converted from their initial .webm format to .mp4 and organized by prompt type using the H.264 codec with the libx264 library (this final video conversion is not performed in version 2.0 in order to minimize re-encoding). While audio/video recordings for the first release, denoted by v1.x on Hugging Face (https://huggingface.co/datasets/faridlab/deepspeak_v1_1), were converted from their initial .webm format to .mp4 using the H.264 codec with the libx264 library, an FFmpeg “copy” command was used in the second release, denoted by v2.x on Hugging Face (https://huggingface.co/datasets/faridlab/deepspeak_v2), to avoid re-encoding and preserve recording quality. The tool output a unique identifier per participant recording that was later used to match recordings to survey responses. To generate video deepfakes, all source videos were re-encoded using FFmpeg with the H.264 codec at a constant rate factor of 18. The encoding present was set to slow, and the original audio stream was preserved without re-encoding.

C.4. Real Video Validation

Participants were given written and visual instructions to allow them to practice recording themselves and test their hardware. Participants were asked to adhere to a series of recording conditions intended to improve consistency within the overall dataset, including positioning themselves centrally within their web camera frame, sitting in a well lit room, not allowing other faces or people to be present in the frame, and minimizing background noise. We manually removed any invalid responses from the final dataset. This included ensuring that only a single person was visible in the video, the scene was reasonably well lit, and each submitted video clip contained a valid audio and visual track. A full list of these conditions is shown in the survey screenshots in Appendix P.

D. Video Generation Methods

D.1. Face-Swap

FaceFusion The first face-swap configuration invokes the FaceFusion [43] library with its default parameters. FaceFusion, built on InsightFace [41], localizes the face in the original video, performs 3D landmark estimation of the face, and

Video Recorder App

Please use this webpage to record yourself responding to the prompts given to you in the study.

A preview of your camera can be seen below.

- Launch the recording window by clicking 'launch recorder' in the preview below.
- In the recording window, click the record button when you are ready to record.
- To stop the recording, press the stop recording button.

Once you have completed recording:

- A 'recording identifier' will be provided below this text in blue. Copy this back into your survey response.
- You can re-record yourself as many times as you wish, but only paste the **final** identifier into the survey response.
- There will be an option to download and playback the video to check it worked properly.

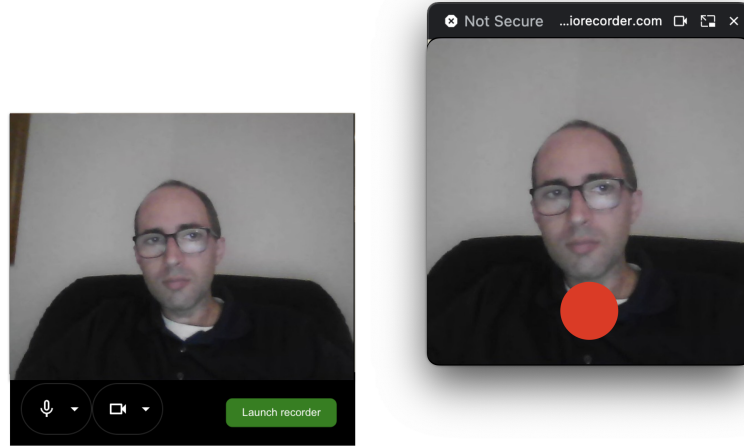


Figure 4. A screenshot of the custom-built recording tool used for real participant data collection. Participants were directed from the data collection survey to the URL hosting the tool. The landing page contained brief instructions for participants to use the recording functionality. Once recording was stopped, a unique identifier was shown to the participant pertaining to that specific recording. This was then input into the survey by the participant, and was used to match responses in the survey to recordings stored by the tool.

superimposes the appropriately transformed matched face. This is followed by a set of post-processing steps to hide edge artifacts and improve temporal consistency.

FaceFusion + GAN: The second face-swap configuration extends the above FaceFusion configuration by enhancing each generated frame using the CodeFormer GAN [64]. This model corrects rendering discrepancies—mostly misshapen teeth and lips—and increases the overall photorealism of the face.

FaceFusion Live: The third face-swap configuration wraps the same FaceFusion configuration in a simulated live input streaming environment. Because our hardware cannot generate video frames in real-time (24 to 30 frames per second (fps)), the effective frame-rate of the generated video is decreased to approximately 12 fps.

INSwapper: The first face-swap configuration invokes the INSwapper-128 model [51] with default parameters. This model leverages InsightFace [41] to localize 3D facial landmarks and superimpose a transformed face of the matched identity onto the source identity video. This version is most similar to FaceFusion in version 1.0.

INSwapper + CodeFormer: The second face-swap configuration extends the above INSwapper configuration by enhancing each frame of the generated deepfake using CodeFormer [63]. This model was trained to fix any misshapen teeth or lips and improve the overall photorealism of the generated face. This version is most similar to FaceFusion+GAN in version 1.0.

SimSwap: The third face-swap configuration invokes the SimSwap-256 model [8] with default parameters. Unlike INSwapper, which directly modifies the target face based on detected facial landmarks, SimSwap extracts a latent identity representation to guide the generation of new frames.

SimSwap + RestoreFormer: The fourth face-swap configuration extends the above SimSwap configuration by enhancing each frame of the generated deepfake using RestoreFormer++ [53]. This model, trained on degraded photographs, corrects any structural deficiencies and enhances the photorealism of the generated frames.

D.2. Lip-Sync

Wav2Lip: The first lip-sync configuration invokes the Wav2Lip model [38] with its default parameters. Wav2Lip is a neural network trained in a GAN-like generator-discriminator fashion. The generator architecture is a LipGAN [26] trained on the LRS2 dataset [1].

VideoRetalking: The second lip-sync configuration invokes the VideoRetalking pipeline [9] with its default parameters. The input video is passed through three neural networks: (1) a semantic-guided reenactment model, which stabilizes the expression in the video; (2) a lip-sync model, which renders a new mouth and chin area matching the audio; and (3) a face enhancer, which fixes rendering discrepancies. The specific models employed in the first two stages are L-Net and D-Net; the third model is the CodeFormer GAN [64].

Diff2Lip: The first lip-sync configuration invokes the Diff2Lip model [34] with default parameters. After localizing the face in the video, Diff2Lip adds noise over the mouth region and proceeds as an audio-conditioned diffusion model. It was trained on the Voxceleb2 [11] and LRW [59] datasets.

LatentSync: The second lip-sync configuration invokes the LatentSync model [27] with default parameters. While LatentSync is, at its core, a diffusion model similar to Diff2Lip, it differs in two aspects. First, it encodes the audio constraint as Whisper embeddings [40] instead of chunks of audio signal. Second, it uses a noise mask delineated by facial landmarks to exactly match the shape of the face as opposed to a rectangular bounding box. LatentSync was trained on the VoxCeleb2 [11] and HDTF [61] datasets.

D.3. Avatar

LivePortrait: The first avatar configuration invokes the LivePortrait model [20] with its default parameters. LivePortrait is a multi-stage model that first extracts the identity and motion embeddings, which are warped into a joint embedding that is later decoded into pixel space. This model was trained on the VoxCeleb [35], MEAD [52], RAVDESS [32], and AAHQ [29] datasets.

HelloMeme: The second avatar configuration invokes the HelloMeme [60] model with its default parameters. HelloMeme consists of three modules: a reference module, which extracts an identity embedding; a control module, which extracts information about the head and mouth shape; and a diffusion module, which generates the resulting video frames. This model was trained on the CelebV-HQ [57] and VFHQ [57] datasets.

Memo: The third avatar configuration invokes the Memo model [62] with its default parameters. Memo is a diffusion model that combines identity, voice, and emotion embeddings as diffusion constraints. It was trained on a compilation of the HDTF [61], VFHQ [57], CelebV-HQ [65], MultiTalk [46], and MEAD [52] datasets.

E. Audio Generation Methods

Three voice clone providers were used to create AI voice clones of all 500 participants, and generate Text-to-Speech audio using these clones.

Firstly, the ElevenLabs Create Voice endpoint (<https://elevenlabs.io/docs/api-reference/voices/add>) was used for voice clone generation, followed by the Create Speech API (<https://elevenlabs.io/docs/api-reference/text-to-speech/convert>) for speech synthesis. The `eleven_multilingual_v2` model was used, with audio output returned in MP3 format at 44.1 kHz. The API was accessed in April 2024.

Secondly, the PlayAI Instant Voice Cloning endpoint (<https://docs.play.ai/reference/api-create-instant-voice-clone>) was used for voice clone generation. The `pyht` Python package, with `TTSOptions` set to default parameters was used for speech synthesis, with audio output provided in MP3 format, at a sampling rate of 24 kHz.

Finally, the Speechify v1 Voices API (<https://docs.sws.speechify.com/v1/api-reference/api-reference/tts/voices/create>) was used for voice clone generation. The Speech API endpoint (<https://docs.sws.speechify.com/v1/api-reference/api-reference/tts/audio/speech>) was used for speech synthesis. The `simba-english` model was used, with output audio returned in WAV format at 48 kHz.

API calls for version 1 were made in April 2024 (ElevenLabs only), and version 2 calls were made in November 2024.

F. Video Validation

While failure types 3, 4, and 5 are general deepfake engine errors that are not flagged by the engines themselves, we found that failure types 1 and 2 often stem from certain features (or lack thereof) in the input. To achieve good performance with face-swap deepfakes, for example, the inputted image needs to show the individual facing the camera, with open eyes and a closed mouth. The same applies to the first frame of videos used as input to avatar deepfakes.

The suite of input and output detectors to filter undesired features performs the following validation types:

Input Validation. Before an image is used as input to a face-swap deepfake engine and before the first frame of a video is used for an avatar-based deepfake, the following criteria are validated. If any of the following criteria are not satisfied, a different image or video is chosen: (1) The participant’s face is fully visible and positioned in the center of the frame, facing the camera. This is established by verifying that there are no hand occlusions over the participant’s face and the overall position and orientation of the participant; (2) The participant’s eyes are open. This is established based on the distance of the X and Y landmark coordinates; or (3) The participant’s mouth is open. This is established based on the distance of the upper and lower lip landmark coordinates.

Output Validation. For each generated video, the following features are validated. If any of these criteria are not satisfied, the video is dropped from the dataset: (1) (*Lip-sync only*) The distance between spectrograms of the original video and the target audio are above a threshold; (2) (*Face-swap only*) The protagonist’s face is more similar to the face of the target identity over the original identity. This is validated by the cosine distance of CLIP embeddings and Structural similarity index measure (SSIM) of faces cropped using MediaPipe [33]; or (3) No frames of the video are fully black.

G. Video Statistics

Table 6. A breakdown of the total size (gigabytes (GB), number of files (N), and length in hours (hrs)) of each version of the DeepSpeak dataset.

Version	Total			Train		Test	
	size (GB)	size (N)	size (hrs)	real (N [hrs])	fake (N [hrs])	real (N [hrs])	fake (N [hrs])
v1	46	13,463	48.0	5,251 [16.8]	5,299 [21.0]	1,416 [4.4]	1,497 [5.8]
v2	124	16,585	52.7	7,513 [23.6]	5,793 [18.6]	1,863 [5.8]	1,416 [4.6]
Total	170	30,048	100.7	12,764 [40.4]	11,092 [39.6]	3,279 [10.2]	2,913 [10.4]

H. Deepfake Video Frame Examples

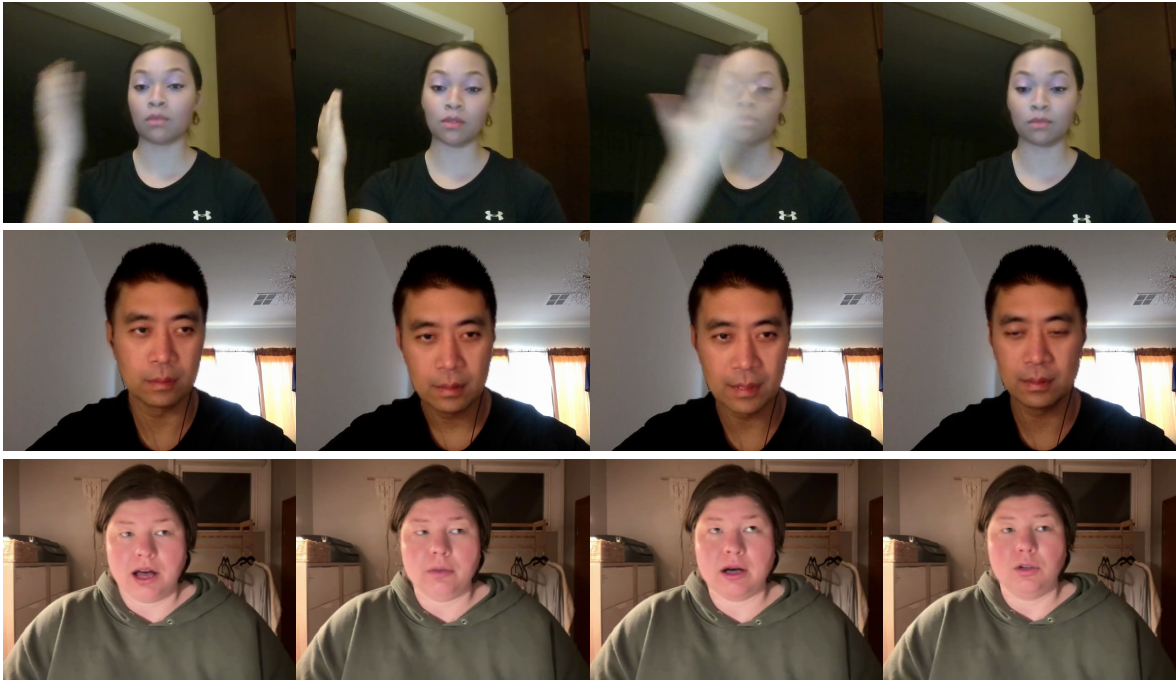


Figure 5. Video frames sampled from three representative examples of **face-swap** deepfakes generated using **FaceFusion**.



Figure 6. Video frames sampled from three representative examples of **face-swap** deepfakes generated using **FaceFusion + GAN**.



Figure 7. Video frames sampled from three representative examples of **face-swap** deepfakes generated using **FaceFusion Live**.

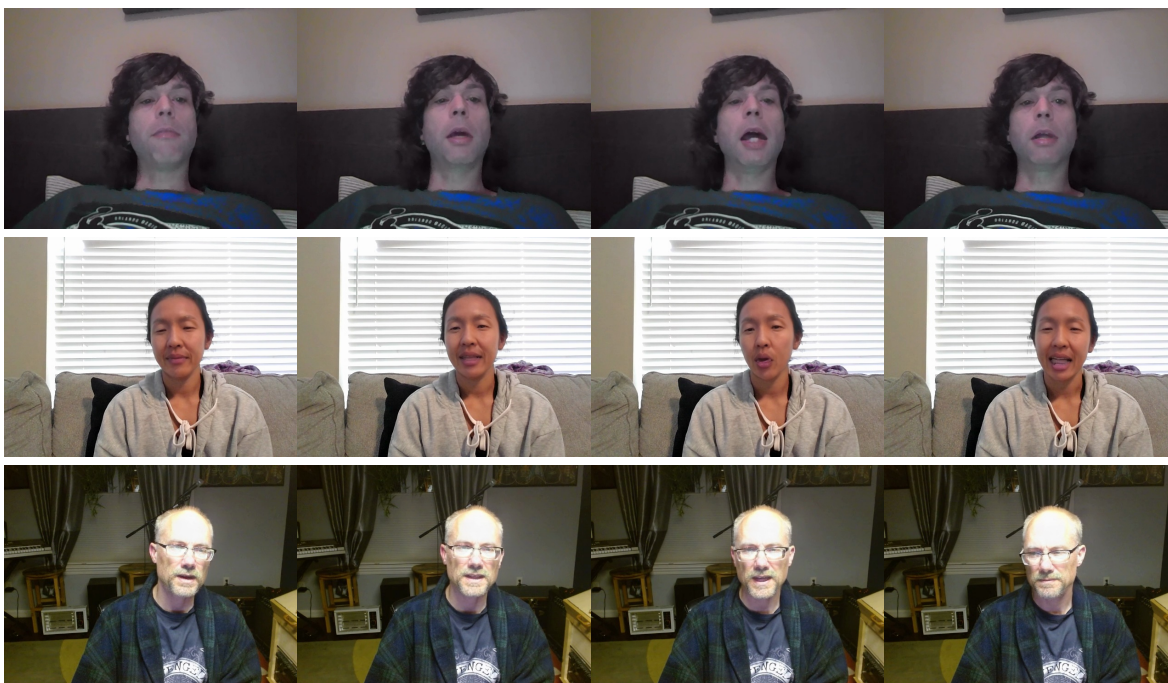


Figure 8. Video frames sampled from three representative examples of **face-swap** deepfakes generated using **INSwapper**.

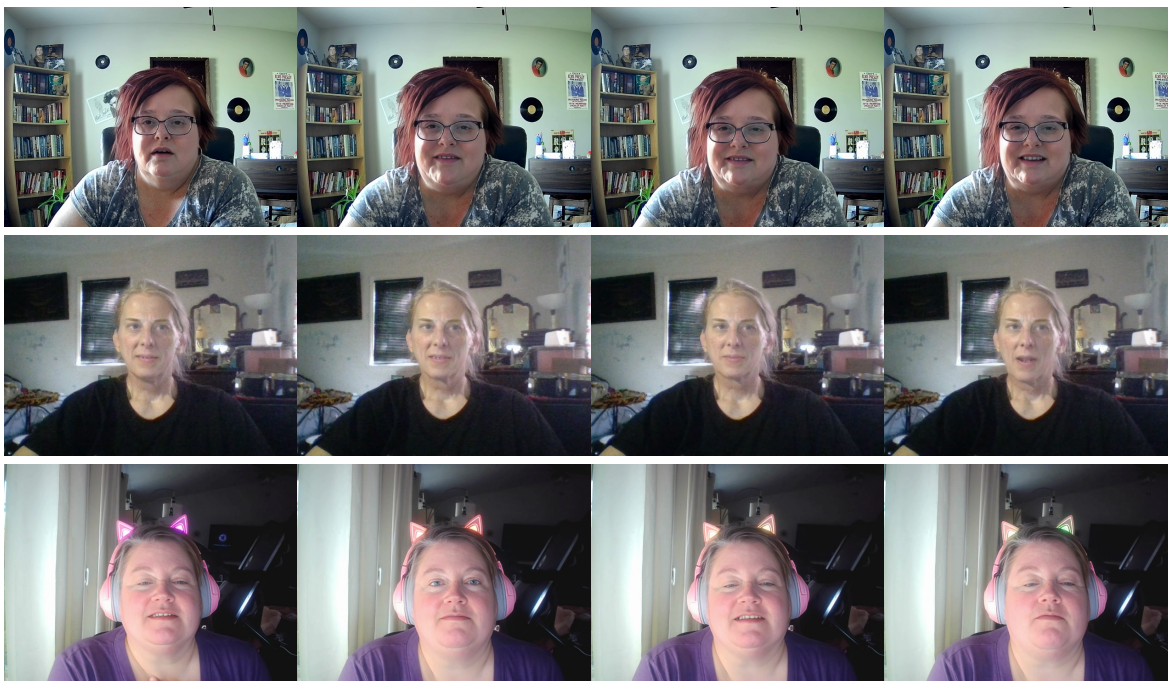


Figure 9. Video frames sampled from three representative examples of **face-swap** deepfakes generated using **INSwapper + CodeFormer**.

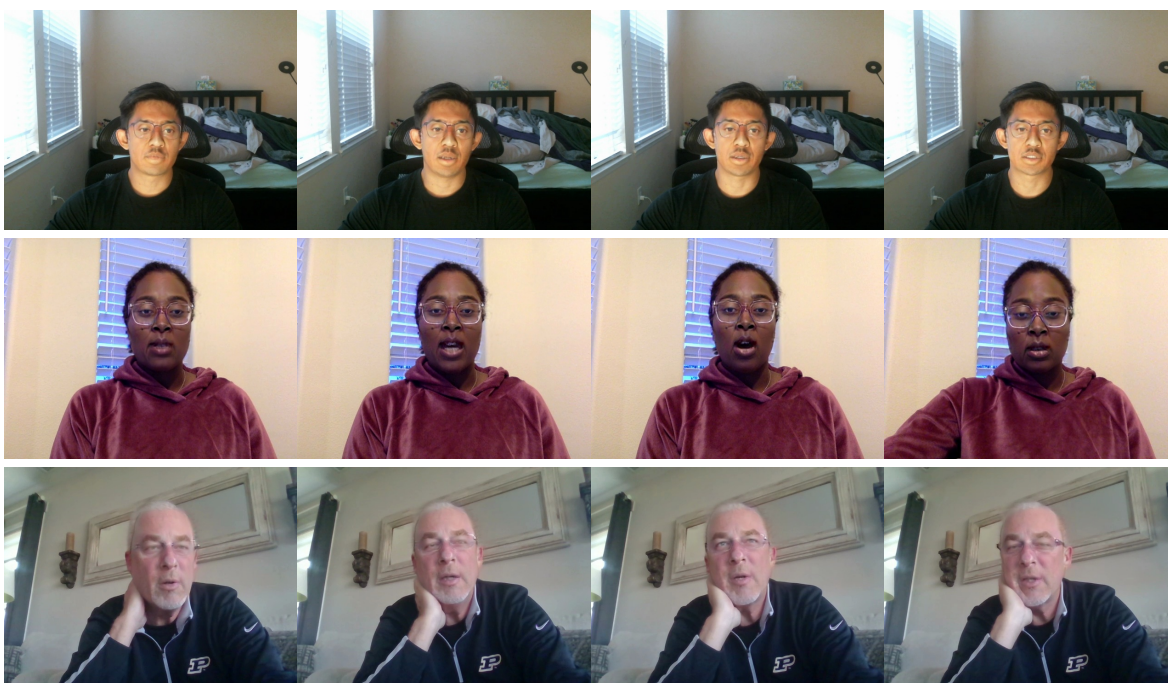


Figure 10. Video frames sampled from three representative examples of **face-swap** deepfakes generated using **SimSwap**.

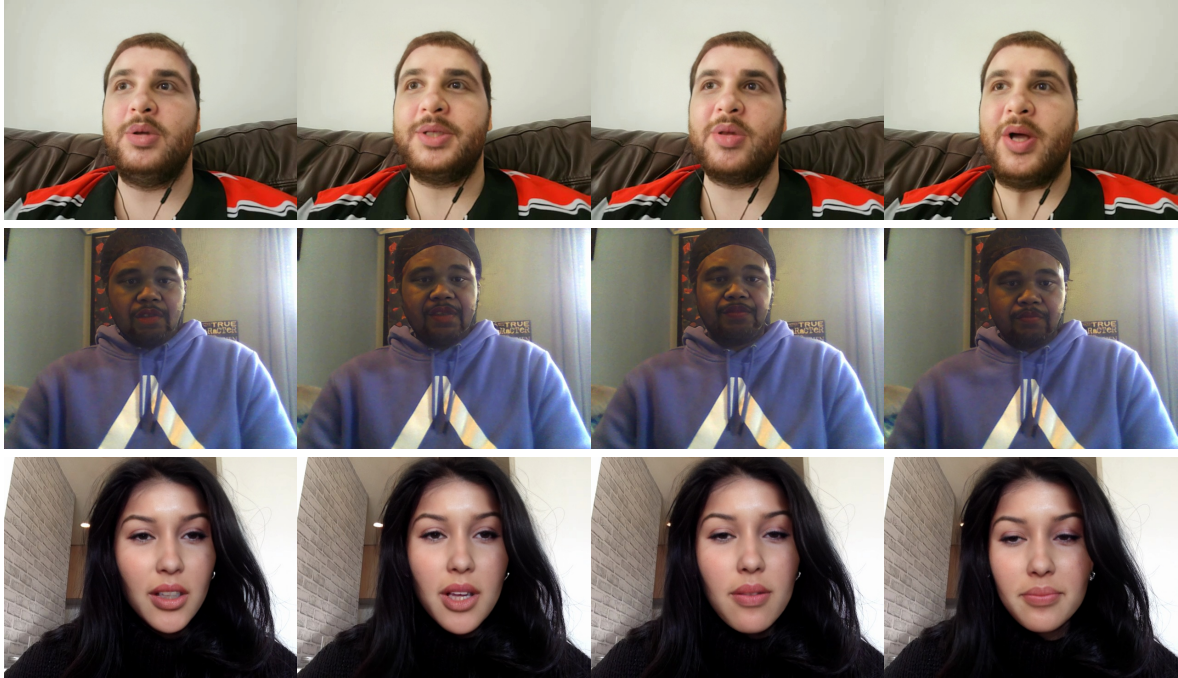


Figure 11. Video frames sampled from three representative examples of **face-swap** deepfakes generated using **SimSwap + RestoreFormer**.



Figure 12. Video frames sampled from three representative examples of **lip-sync** deepfakes generated using **Wav2Lip**.



Figure 13. Video frames sampled from three representative examples of **lip-sync** deepfakes generated using **VideoRetalking**.

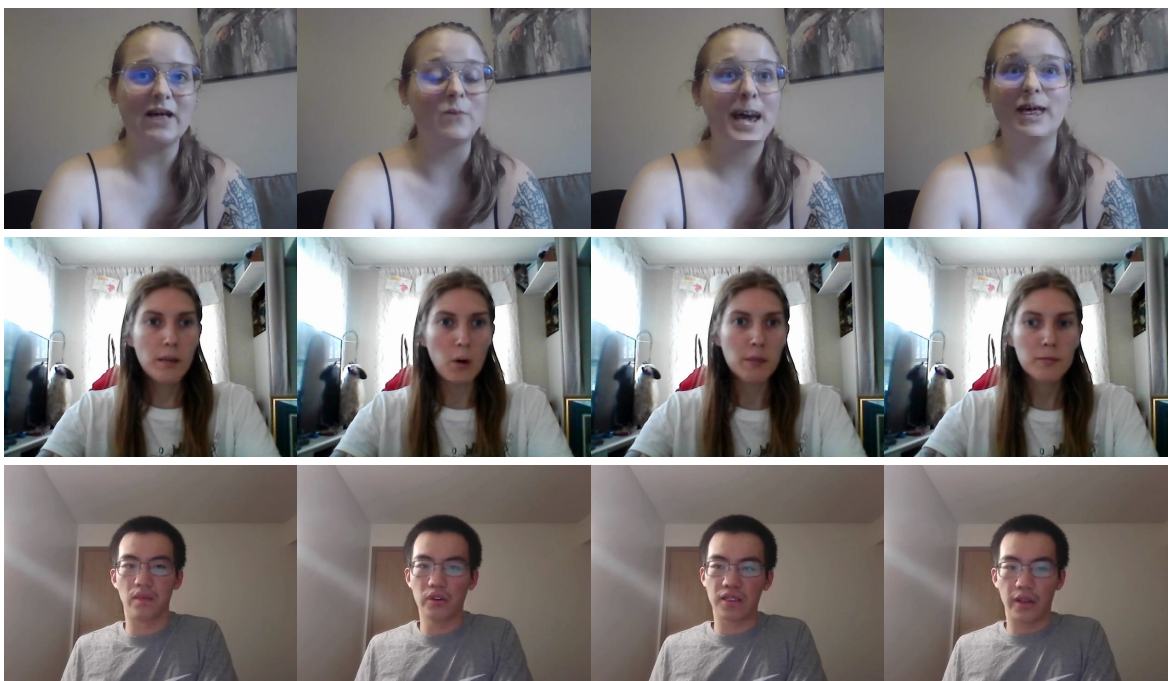


Figure 14. Video frames sampled from three representative examples of **lip-sync** deepfakes generated using **Diff2Lip**.

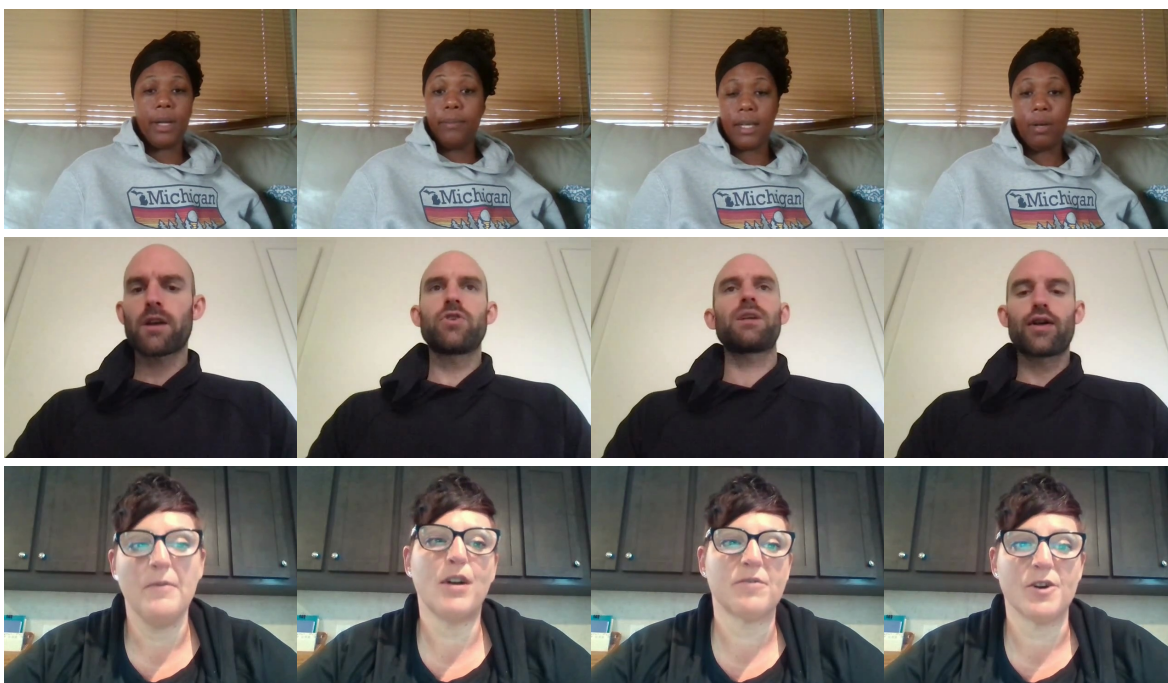


Figure 15. Video frames sampled from three representative examples of **lip-sync** deepfakes generated using **LatentSync**.

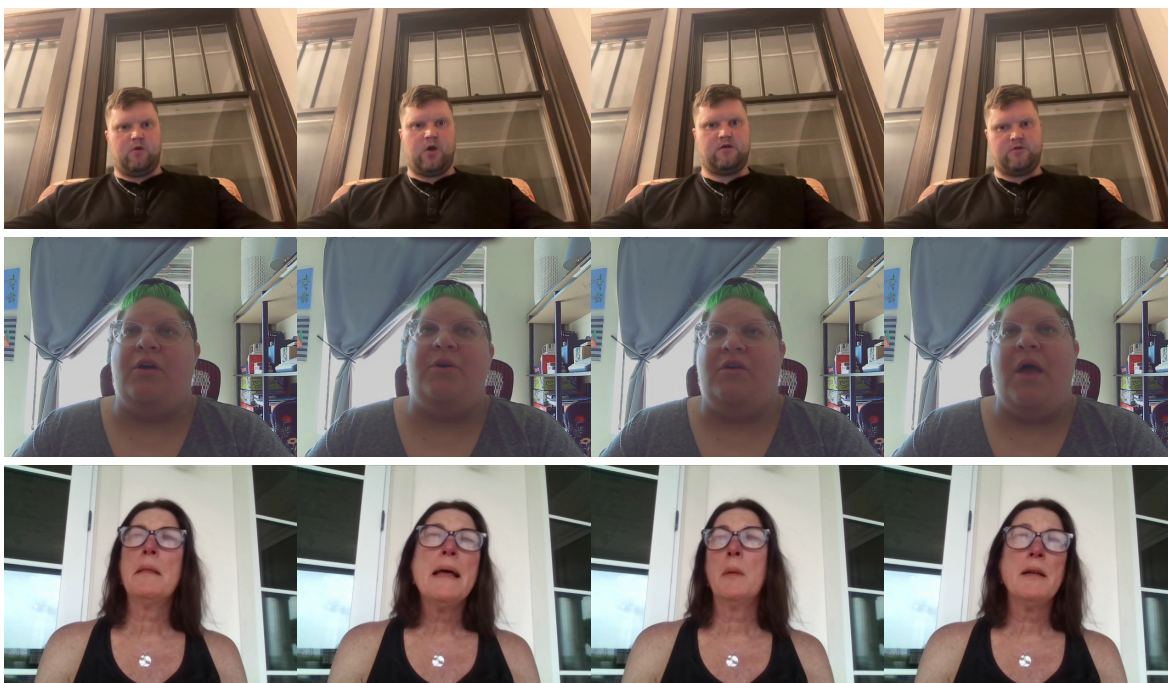


Figure 16. Video frames sampled from three representative examples of **avatar** deepfakes generated using **LivePortrait**.

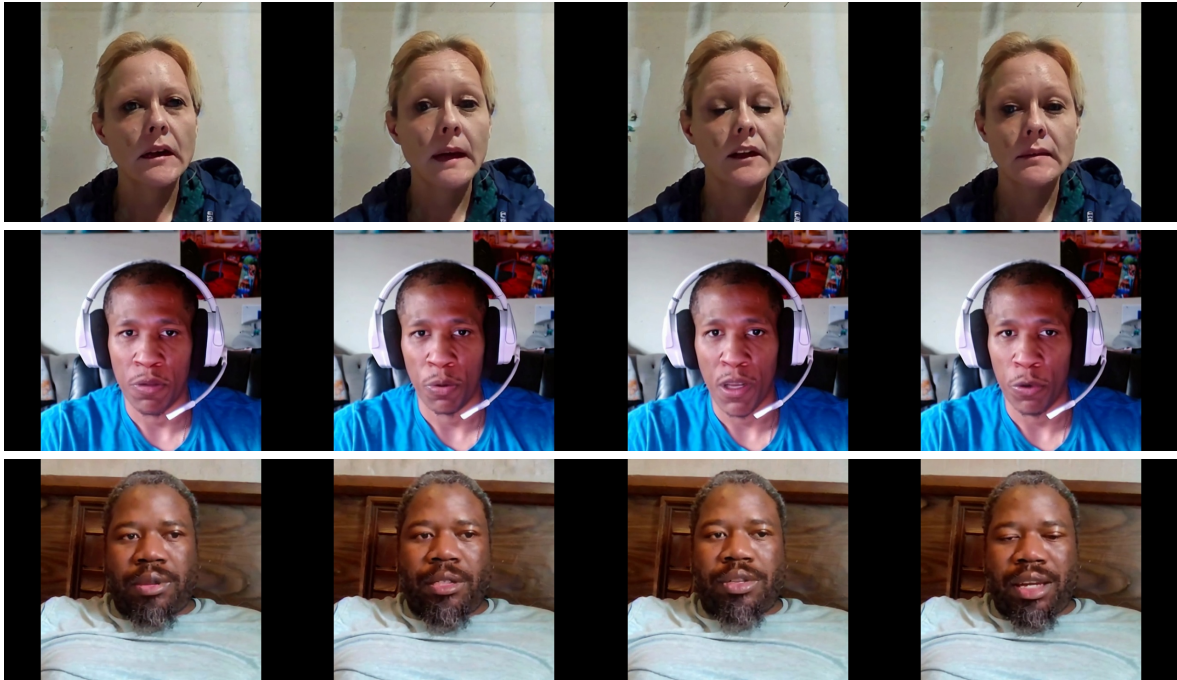


Figure 17. Video frames sampled from three representative examples of **avatar** deepfakes generated using **HelloMeme**.

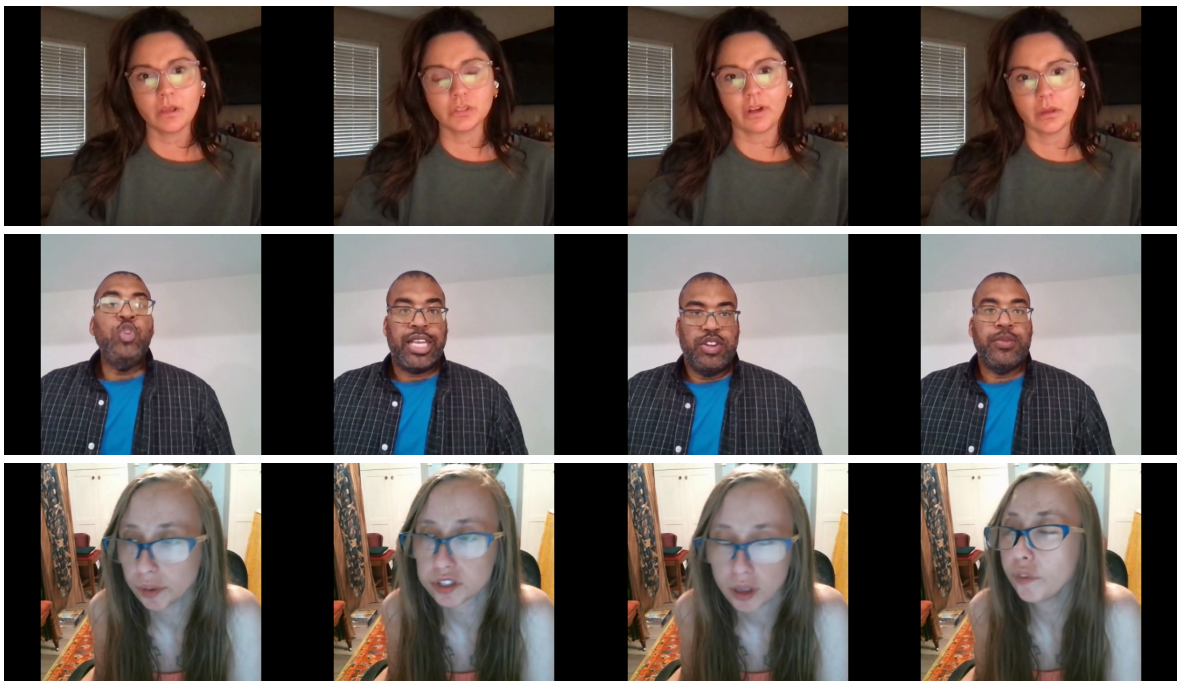


Figure 18. Video frames sampled from three representative examples of **avatar** deepfakes generated using **Memo**.

I. Visual Identity Pairing Examples



Figure 19. Additional examples of matched identities from the first release of DeepSpeak data (v1.x). Part 1/2.



Figure 20. Additional examples of matched identities from the first release of DeepSpeak data (v1.x). Part 2/2.



Figure 21. Additional examples of matched identities from the first release of DeepSpeak data (v2.0). Part 1/2.



Figure 22. Additional examples of matched identities from the first release of DeepSpeak data (v2.x). Part 2/2.

J. Audio Identity Matching

J.1. Audio Identity Matching Approach

Audio identities can be matched in a similar way to visual identities. While not currently implemented in the DeepSpeak dataset, we have found this approach to be useful in other work [4] and include it here to support future research on audio identity matching, and to enable future versions of the dataset where audio identities can be swapped alongside visual identities.

Each identity can be represented by an embedding generated from one input audio per participant. For this task, we select TitaNet-L, a neural model primarily used for speaker recognition. It employs separable convolutions with Squeeze-and-Excitation and pooling layers to map variable-length utterances to 192-dimensional embeddings. While alternative embedding systems exist, we opt for TitaNet-L for two main reasons: (1) its speaker recognition architecture enables identity-based matching, and (2) it perceptually performed well in terms of qualitative results. TitaNet, alongside other embedding systems, is explored for the purposes of fake detection in 6.2. The same short scripted audio was used for all identities. As with the visual CLIP embeddings, comparing t-SNE representations of self-reported demographic information reveals that TitaNet embeddings cluster around gender.

Each identity can then be matched according to the same matching process outlined in Section 3. The matched pairs are provided in Appendix J.2.

J.2. Audio Identity Pairs

Table 7. Audio identity pairs generated by taking the minimal cosine distance between vectors comprising 192-dimensional TitaNet embedding values.

identity_1	identity_2
152	491
10	32
393	404
43	285
236	333
364	371
135	255
130	455
151	268
105	466
120	374
132	294
74	362
103	112
18	219
147	467
162	192
109	332
157	405
241	437

identity_1	identity_2
104	180
35	174
53	445
70	377
51	244
281	501
44	189
270	359
357	413
69	273
209	350
200	277
345	366
65	478
11	344
62	66
153	399
54	226
129	278
95	114

identity_1	identity_2
52	386
90	227
170	410
19	265
4	338
203	476
186	261
56	86
61	312
206	330
111	428
41	214
423	482
418	459
88	149
98	381
115	353
20	154
17	117
246	297

K. Experimental Setup: Data Preprocessing

K.1. Video Deepfake Detection

K.1.1. FreqNet

Original Dataset. A subset of the GAN-based deepfake dataset compiled by FreqNet’s authors, downloaded using the official script (https://github.com/chuangchuangtan/FreqNet-DeepfakeDetection/blob/main/download_dataset.sh), was used for evaluation. This subset comprised AttGAN, RelGAN, and S3GAN samples.

DeepSpeak. The DeepSpeak videos were split into frames and analyzed individually, consistent with FreqNet’s pre-processing.

K.1.2. GenConViT ED/VAE

Original Dataset. CelebDF-2, downloaded from its Kaggle clone (<https://www.kaggle.com/datasets/reubensuju/celeb-df-v2>), was used for evaluation.

DeepSpeak. The DeepSpeak videos were cropped to the protagonist’s face, consistent with GenConViT’s pre-processing, as described in the paper. Since the authors did not release a pre-processing script as a part of their official code release, we implemented this script according to its description in the paper. This script is available at [anonymized](#).

K.1.3. LipFD

Original Dataset. AVLips, downloaded from <https://github.com/AaronComo/LipFD?tab=readme-ov-file#-avlips-a-high-quality-audio-visual-dataset-for-lipsync-detection>, was used for evaluation.

DeepSpeak. The DeepSpeak videos were pre-processed into grids with a sequence of five crops of the protagonist’s face at the bottom, with a matching spectrogram at the top. This pre-processing was performed using the official script released by LipFD’s authors (<https://github.com/AaronComo/LipFD/blob/main/preprocess.py>).

K.2. Audio Deepfake Detection

K.2.1. Raw Waveform Models

Original Dataset AASIST was trained by its authors using the training subset of the ASVSpooF dataset[23]. The authors provide two pretrained benchmark models (RawNet2 and RawGAT-ST) implemented as .py files associate with weights stored in .pth format, all generated using ASVSpooF training data <https://github.com/clovaai/aasist>.

DeepSpeak. For DeepSpeak, the audios samples were used in their full duration as raw waveforms. These were extracted from the original .mp4 files, saved as .wav format, and resampled to 16kHz.

K.2.2. Embedding-based Models

TitaNet. The original TitaNet + classifier model was trained on the TIMIT-ElevenLabs dataset. While the exact embedding extraction and classifier building code are not publicly available, we re-implemented the approach as described in the original work and retrained it on the same dataset. Embeddings were extracted from full-duration audio waveforms, resulting in 192-dimensional vectors per sample which were stored in CSV format and used as input to the downstream classifiers.

Wav2Vec2-xlsr. For “pretrained” models, we trained the Wav2Vec2-xlsr classifier on embeddings extracted from ASVSpooF audios, as we did not find any widely used or well-documented pretrained implementations available. Embeddings were extracted from full-duration audio waveforms, resulting in 1024-dimensional vectors per sample which were stored in CSV format and used as input to the downstream classifiers. Embeddings from the DeepSpeak data were extracted in the same way.

LAION-CLAP. We trained the LAION-CLAP classifier on ASVSpooF, following the same procedure as used for the Wav2Vec2-xlsr embeddings. Embeddings were extracted from full-duration audio waveforms, resulting in 512-dimensional vectors per sample which were stored in CSV format and used as input to the downstream classifiers. Embeddings from the DeepSpeak data were extracted in the same way.

L. Experimental Setup: Hyperparameters

L.1. Video Deepfake Detection

Default hyperparameters values were used when possible. These were taken from the official code repositories: <https://github.com/chuangchuangtan/FreqNet-DeepfakeDetection> for FreqNet, <https://github.com/erprogs/GenConViT> for GenConViT, and <https://github.com/AaronComo/LipFD> for LipFD. Exceptions to default hyperparameter values are marked with *. These include the learning rate and weight decay parameters, which sometimes had to be adapted for the models to learn and converge successfully (a small search over neighboring magnitudes of the default value was performed). We also had to change the batch size to accommodate our compute resources.

L.2. Audio Deepfake Detection

Details of the hyperparameters used in training the raw waveform models are included in the next subsection. For the embedding-based models, data was subset on a 80/20% training/testing split, with training data balanced through downsampling

Parameter	Full Training			Fine-tuning		
	FreqNet	GenConViT	LipFD	FreqNet	GenConViT	LipFD
Training epochs	10	10	10	10	10	10
Batch size	32	16	10	32	16	10
Frame size	256x256	224x224	500x200	256x256	224x224	500x200
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate (LR)	1e-4	1e-4	1e-5	1e-4	1e-4	1e-5
LR scheduler gamma	N/A	1e-1	N/A	N/A	1e-1	N/A
LR scheduler step size	N/A	15	N/A	N/A	15	N/A
Weight decay	N/A	1e-4	1e-4	N/A	1e-4	1e-4

Table 8. Hyperparameters used for full training and fine-tuning of video deepfake detection models.

the larger class to match the number of audios in the smaller class (in all cases, the “real” class was larger than the “fake” class due to the fact that only lip-sync deepfakes in DeepSpeak contain fake audio). However, all testing data was used for inference.

Logistic regression and random forest classifiers were used as the linear and non-linear classification models for real and fake. The following parameters were used by default for both models across all datasets:

1. LogisticRegression: max iterations = 1000, with all other parameters as per the defaults in the Scikit-learn LogisticRegression model.
2. RandomForestClassifier: number of estimators = 100, with all other parameters as per the defaults in the Scikit-learn RandomForestClassifier model.

L.3. Audio Raw Waveform Pretrained Model Configurations

Listing 1. AASIST Full Configuration

```
{
  "database_path": "./LA/",
  "asv_score_path": "ASVspoof2019_LA_asv_scores/ASVspoof2019.LA.asv.eval.gi.trl.scores.txt",
  "model_path": "./models/weights/AASIST.pth",
  "batch_size": 24,
  "num_epochs": 100,
  "loss": "CCE",
  "track": "LA",
  "eval_all_best": "True",
  "eval_output": "eval_scores_using_best_dev_model.txt",
  "cudnn_deterministic_toggle": "True",
  "cudnn_benchmark_toggle": "False",
  "model_config": {
    "architecture": "AASIST",
    "nb_samp": 64600,
    "first_conv": 128,
    "filtfs": [70, [1, 32], [32, 32], [32, 64], [64, 64]],
    "gat_dims": [64, 32],
    "pool_ratios": [0.5, 0.7, 0.5, 0.5],
    "temperatures": [2.0, 2.0, 100.0, 100.0]
  },
  "optim_config": {
    "optimizer": "adam",
    "amsgrad": "False",
    "base_lr": 0.0001,
    "lr_min": 0.000005,
    "betas": [0.9, 0.999],
    "weight_decay": 0.0001,
    "scheduler": "cosine"
  }
}
```

Listing 2. RawNet2Spoof Full Configuration

```
{
  "database_path": "./LA/",
  "asv_score_path": "ASVspoof2019_LA_asv_scores/ASVspoof2019.LA.asv.eval.gi.tr1.scores.txt",
  "model_path": "/home1/irteam/jeeweon/git/AsvSpoofDetection/exp_result/LAmodelRawNet2Spoof_ep100_bs32_lr0.0001/weights/best.pth",
  "batch_size": 32,
  "lr": 0.0001,
  "weight_decay": 0.0001,
  "num_epochs": 100,
  "loss": "CCE",
  "track": "LA",
  "eval_output": "eval_scores_using_best_dev_model.txt",
  "cudnn_deterministic_toggle": "True",
  "cudnn_benchmark_toggle": "False",
  "model_config": {
    "architecture": "RawNet2Spoof",
    "nb_samp": 64600,
    "first_conv": 1024,
    "in_channels": 1,
    "filtfs": [20, [20, 20], [20, 128], [128, 128]],
    "blocks": [2, 4],
    "nb_fc_node": 1024,
    "gru_node": 1024,
    "nb_gru_layer": 3,
    "nb_classes": 2
  },
  "optim_config": {
    "optimizer": "adam",
    "amsgrad": "False",
    "base_lr": 0.0001,
    "lr_min": 0.000005,
    "betas": [0.9, 0.999],
    "weight_decay": 0.0001,
    "scheduler": "cosine"
  }
}
```


Listing 3. RawNetGatSpoofST Full Configuration

```
{
  "database_path": "./LA/",
  "asv_score_path": "ASVspoof2019_LA_asv_scores/ASVspoof2019.LA.asv.eval.gi.trl.scores.txt",
  "model_path": "/home1/irteam/jeeweon/git/AsvSpoofDetection/exp_result/LAmodelRawNetGatSpoofST_ep100_bs24_lr0.0001/weights/epoch_12.pth",
  "batch_size": 24,
  "num_epochs": 100,
  "loss": "CCE",
  "track": "LA",
  "eval_output": "eval_scores_using_best_dev_model.txt",
  "cudnn_deterministic_toggle": "True",
  "cudnn_benchmark_toggle": "False",
  "model_config": {
    "architecture": "RawNetGatSpoofST",
    "nb_samp": 64600,
    "first_conv": 128,
    "filtfs": [70, [1, 32], [32, 32], [32, 64], [64, 64]]
  },
  "optim_config": {
    "optimizer": "adam",
    "amsgrad": "False",
    "base_lr": 0.0001,
    "lr_min": 0.000005,
    "betas": [0.9, 0.999],
    "weight_decay": 0.0001,
    "scheduler": "cosine"
  }
}
```

L.4. Audio Raw Waveform Model Configurations for Training on DeepSpeak

The model configurations used for training AASIST, RawNet2 and RawGAT-ST on DeepSpeak data were the same as those for the original pretrained models created by the AASIST authors (using ASVspoof) as detailed in the previous subsection, except with epochs reduced from 100 to 50 for training efficiency.

M. Safeguards against Misuse

While deepfake detection technology is broadly beneficial for enhancing the security and safety of individuals, organizations, and societies against harms such as scams, fraud, and disinformation, we acknowledge the potential for misuse of the dataset and outline our safeguards against this below.

We make the data available under license via Hugging Face, where metadata and documentation are publicly visible. However, access to the data itself is restricted: users must request access and briefly describe their intended use. Only projects that aim to improve defenses against deepfakes or support reproducibility studies will be granted access.

By way of safeguards for participants in the data collection, we explicitly obtained consent from all users and informed them that their recordings — but no other personally identifiable information (PII) — would be included in a publicly available dataset. Participants were instructed not to include other individuals in their recordings. While we report an overview of the participants in this paper, we do not release individual-level demographic information.

N. Licensing

This section lists licensing terms for all assets employed in this work at the time of submission (May 2025). We refer the reader to the respective publications or code repositories for the most up-to-date licensing terms.

N.1. DeepSpeak Dataset

This is a custom asset associated with this paper. Licensing is provided to qualifying academic institutions at no cost under licensing terms available at . Deepfakes included in DeepSpeak were generated with the following third-party assets (deepfake engines).

FaceFusion is provided under the OpenRAIL-AS license.

INSwapper's code repository does not specify a license.

CodeFormer is provided under a custom license posted at <https://github.com/sczhou/CodeFormer?tab=License-1-ov-file>.

SimSwap is provided under the Creative Commons Attribution-NonCommercial 4.0 International license.

RestoreFormer is provided under the Apache v2.0 license.

Wav2Lip's code repository does not specify a license.

VideoRetalking is provided under the Apache v2.0 license.

Diff2Lip is provided under the Creative Commons Attribution-NonCommercial 4.0 International license.

LatentSync is provided under the Apache v2.0 license.

LivePortrait is provided under the MIT license.

HelloMeme is provided under the MIT license.

Memo is provided under the Apache v2.0 license.

N.2. Video Deepfake Detection Experiments

The following are third-party assets (model code and datasets) used for the video deepfake detection experiments.

FreqNet's code repository does not specify a license.

GenConViT is provided under the GNU General Public License v3.0.

LipFD's code repository does not specify a license.

Celeb-DF 2's data repository does not specify a license.

AVLips's data repository does not specify a license.

N.3. Audio Deepfake Detection Experiments

Both ElevenLabs and PlayAI agreed to grant our team complimentary research access to their commercial voice cloning and Text-to-Speech APIs. For Speechify, the paid Premium commercial tier was used.

ElevenLabs's website specifies a commercial license is granted for use of audio created with the API, in accordance with their Terms of Service (<https://help.elevenlabs.io/hc/en-us/articles/13313564601361-Can-I-publish-the-content-I-generate-on-the-platform>).

PlayAI's Terms of Service grant users, solely for commercial use, all right, title and interest in and to content generated by the Service based on your User Content ("Output"), subject to any Third Party Terms which may apply to such Output (<https://play.ai/terms>).

Speechify's Terms of Service specify a commercial license is granted for audios generated using the paid subscription tier (<https://speechify.com/studio-terms/?srsltid=AfmBOopH.aautZGkvdGqxbBAAl-d-shNU94UHq8v-xu0wUz1coqYaHJqJ>).

O. Compute Resources

We conducted our experiments on single-node NVIDIA A100 GPU machines. All reported runtimes correspond to this setup, assuming no competing processes. In total, producing the deepfakes and conducting baseline experiments required approximately 1,700 hours of GPU time, excluding environment setup or troubleshooting.

O.1. DeepSpeak Dataset

Producing the DeepSpeak video deepfakes required approximately eight weeks of cumulative GPU time, including quality validation and filtering. This process was distributed across multiple GPU machines, with each machine generating a subset of the dataset.

Producing the DeepSpeak audio deepfakes involved approximately one week of GPU time, including audio preprocessing, voice cloning, and Text-To-Speech generation using relevant API calls. The entire process was executed on a single GPU machine.

O.2. Video Deepfake Detection Experiments

FreqNet Training or fine-tuning for 10 epochs on DeepSpeak takes approximately ten hours. Evaluating the resulting model on testing sets of DeepSpeak and the dataset of GAN-generated deepfakes produces by FreqNet's authors takes less than 15 minutes. Given one model training, one model fine-tuning, and three evaluation passes, experiments with this architecture required approximately 21 hours of GPU time.

GenConViT Training or fine-tuning for 10 epochs on DeepSpeak takes approximately seven hours. Evaluating the resulting model on testing sets of DeepSpeak and Deleb-DF 2 takes less than 15 minutes. Given two model trainings, two model fine-tunings, and five evaluation passes, experiments with this architecture required approximately 29 hours of GPU time.

LipFD Training or fine-tuning for 10 epochs on DeepSpeak takes approximately 30 hours. Evaluating the resulting model on testing sets of DeepSpeak and AVLips takes under one hour. Given one model training, one model fine-tuning, and three evaluation passes, experiments with this architecture required approximately 63 hours of GPU time.

O.3. Audio Deepfake Detection Experiments

Raw waveform models Training each of AASIST, RawGat-ST and RawNet2 on DeepSpeak data from scratch takes approximately 8 hours. Evaluating the results each model for inference on DeepSpeak takes under two hours.

Embedding-based models Embedding generation takes approximately 4 hours per model (TitaNet-L, Wav2Vec2-xlsr and LAION-CLAP) on DeepSpeak training data and approximately 6 hours per model on ASVSpooof training data. Classifier training (logistic regression and random forest) take less than one hour per embedding type on both ASVSpooof and DeepSpeak training data. Inference for both models per testing set of each dataset takes less than one hour.

P. Survey Materials

Video Recording Study

In the following survey, you will be asked to record yourself using your webcam whilst responding to **30 short prompts**. The survey will be divided into **3 sections** and should take you **about 40 minutes** to complete.

You will then be given a **completion ID** that you must copy in order to complete the survey task.

Once completed, your recordings will be validated. Payment will be provided once your submission has been approved. **Empty recordings, heavily noisy audios or incorrect video actions will not be accepted.**

Please ensure that you are in a **somewhat quiet location** with a **plain background** and that the room is **well lit**. It is important that **no other faces are visible**, even in paintings. Some background noise is acceptable, provided you can still clearly hear your own voice when you replay the video.

Finally, this dataset will be used for research purposes for detecting deepfakes. Please note that your recordings will be made public in a dataset, but no other identifying information will be shared outside of our research group.

Please select the option below to consent to participate in this study.

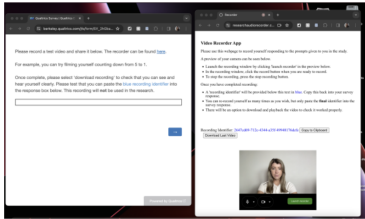
☐ I consent to participate

→

Figure 23. Screenshot of the introduction and consent page of the data collection study.

You will be asked to record your voice using our **video recording tool**. This can be found [here \(https://\[redacted\]\)](https://[redacted]). Please note that this tool only works on **Google Chrome browser**. The tool will try to record in high definition; if you have issues with loading the tool, please check the strength of your internet connection, refresh the page and wait a few moments.

A series of prompts will be shown to you. Please use the tool to record yourself responding to each. You will most likely need two separate browser windows open side by side to enable this, as shown below:



You can **download the video** to ensure that the recording worked successfully. You can double-click the downloaded .webm file to open and watch it. **It is essential that both your audio and video are working correctly; otherwise, your submission will unfortunately be rejected.**

Once you are satisfied with the recording, paste the **blue recording identifier** into the response box in this survey, shown on the left above. **You will need to click 'launch recorder' each time that you create a new recording.**

You can re-record each video as many times as you wish. Please only paste the final recording identifier for the video that you are satisfied with.

You will be able to test the recorder in the next question.

Figure 24. Screenshot of the overview and recording instructions page of the data collection study.

Now that the recorder is working, we need to check your recording environment.

Please situate yourself in the middle of the frame as shown below. Your full face should be visible, from forehead to at least your neck, like this:



Please ensure the following necessary conditions are met. If any of these necessary conditions are not met, our software may not be able to validate your recording and you may not be compensated for your time. You can take as much time as you need to prepare your environment.

- The room is quiet with minimal background noise
- The room is well-lit
- There are no other people or faces present (including in art works)
- The background is mostly plain (some shelves, objects, and background view of the room is fine)
- Your web camera is directly facing you straight on (e.g. ensure you are not looking away at a separate monitor)
- You are in the middle of the frame
- Your web camera is on a flat surface and isn't moving
- Your face is fully in the frame (from forehead to neck) as shown above
- Your internet connection is stable

The following conditions are helpful, but not necessary:

- If possible, please do not wear over-the-ear headphones

Please check the box below to continue.

☐ I confirm that I have met all necessary conditions to complete this study

[Next](#)

Figure 25. Screenshot of the environment checks page of the data collection study.

Please record a test video according to the previous conditions and share it below. The recorder can be found [here](#).

For example, you can try filming yourself counting down from 5 to 1.

Once complete, please select 'download recording' to check that you can see and hear yourself clearly.

If you are experiencing issues with the recording process, please note that the tool requires camera hardware with a minimum resolution of 720p and so you may not be able to use the tool successfully if this is not the case.

Please test that you can paste the [blue recording identifier](#) into the response box below.

[Next](#)

Figure 26. Screenshot of the recording test page of the data collection study.

Section 1 of 2

Please record yourself reading aloud the following sentences. Use the [recording tool](#) and paste the blue recording ID in the box once complete.

*The birch canoe slid on the smooth planks.
Glue the sheet to the dark blue background.
It's easy to tell the depth of a well.
These days a chicken leg is a rare dish.
Rice is often served in round bowls.
The juice of lemons makes fine punch.
The box was thrown beside the parked truck.
The hogs were fed chopped corn and garbage.
Four hours of steady work faced us.
A large size in stockings is hard to sell.*

Figure 27. Screenshot of an example prompt from the data collection study.

Q. Prompts

Q.1. Voice Cloning Prompts

Table 9. Voice cloning input prompts, consisting of ten consecutive sentences and one continuous paragraph.

The birch canoe slid on the smooth planks. Glue the sheet to the dark blue background. It's easy to tell the depth of a well. These days a chicken leg is a rare dish. Rice is often served in round bowls. The juice of lemons makes fine punch. The box was thrown beside the parked truck. The hogs were fed chopped corn and garbage. Four hours of steady work faced us. A large size in stockings is hard to sell.
Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Q.2. Standardized Scripted Prompts

Table 10. Standardized scripted prompts, consisting of ten separate sentences.

We apply auditory modeling to computer speech recognition.
He sank back sighing and was soon asleep again.
The mango and the papaya are in a bowl.
Will you tell me why.
That experience holds a lesson for us all in regard to birth control today.
That is what childhood is he told himself.
As a rule part time farmers hire little help.
The new birth is miraculous and mysterious.
The fear of punishment just didn't bother him.
The figure in the corner belched loudly a deep liquid eruption.

Q.3. Unscripted Prompts

Table 11. Unscripted prompts.

Unscripted Prompts
What did you do yesterday?
Please describe an object that you can see from your current position.
What is your favorite type of music and why?
Please describe your ideal way to spend a weekend.

Q.4. Video Action Prompts

Table 12. Video action prompts.

Versions 1.0 and 2.0
Lean forwards towards the camera and back to your starting position.
1. Look down and to the right, and slowly count out loud to three. 2. Look down and to the middle, and slowly count out loud to three. 3. Look down and to the left, and slowly count out loud to three.
1. Look up and to the left, and slowly count out loud to three. 2. Look up and to the middle, and slowly count out loud to three. 3. Look up and to the right, and slowly count out loud to three.
Wave your hand back and forth across your face four times while counting out loud.
Read each question aloud, followed immediately by your answer. 1. What is my favorite food? 2. What is my favorite movie? 3. What did I have for breakfast? 4. Where would I most like to travel and why? 5. If I could time travel, where in the past or future would I like to go?
Version 2.0 only
Pretend to yawn, perhaps covering your mouth with your hand.
Pretend to laugh for between 2 to 5 seconds.
Clap loudly three times, pausing for about 5 seconds between claps. Please make sure your clap makes a loud sound.

Q.5. Example frames from action prompts



Figure 28. Representative examples of different action prompt types.

Q.6. Randomized Scripted Prompts

She had your dark suit in greasy wash water all year.	Her classical performance gained critical acclaim.	The diagnosis was discouraging however he was not overly worried.
Rich looked for spotted hyenas and jaguars on the safari.	The football team coach has a watch thin as a dime.	He found an empty bench opened a newspaper and stretched his legs before him.
The sermon emphasized the need for affirmative action.	You're so preoccupied that you've let your faith grow dim.	Gus saw pine trees and redwoods on his walk through sequoia national forest.
Approach your interview with statuesque composure.	The annoying raccoons slipped into phil's garden every night.	Some tore entirely through the whipsawed post oak.
Markets should become more competitive as consumers become more selective.	Pledge to participate in nevada's aquatic competition.	The advertising verse of plymouth variety store never changes.
The eastern coast is a place for pure pleasure and excitement.	She always could sense the shag end of a woolly day.	These planets were much bigger nearly all capable of holding an atmosphere.
He had not covered a hundred yards before a gun crashed from somewhere behind.	Between meetings he helps the president keep track of delegated matters.	A crab challenged me but a quick stab vanquished him.
Just why anybody should wish to start a riot the executive officer didn't know.	They used an aggressive policeman to flag thoughtless motorists.	The source is known so there is no necessity to remove insecticide residues.
A toothpaste tube should be squeezed from the bottom.	A precision transit is set up so that it is lined with respect to true north.	Two gas lamps were no more than a misleading glow.
Those who are not purists used canned vegetables when making stew.	Unless we send out the whole pie their pieces mean nothing.	The easygoing zoologist relaxed throughout the voyage.
His failure to open the store by eight cost him his job.	Smash lightbulbs and their cash value will diminish to nothing.	Etiquette mandates compliance with existing regulations.
Remember to allow identical twins to enter freely.	She always jokes about too much garlic in his food.	The government sought authorization of his citizenship.
Children can consume many fruit candies in one sitting.	Bob bandaged both wounds with the skill of a doctor.	It gives social guidance and direction and makes for programs of social action.
No signs of these no gross hemorrhage of lungs heart brain or stomach.	Herb's birthday occurs frequently on thanksgiving.	All about him stood tombstones his own sensitive great hands had fashioned.
With the spring rains the flow rose rapidly due to infiltration in open sewers.	Scholastic aptitude is judged by standardized tests.	We could barely see the fjords through the snow flurries.
A concurrent effort is needed to make oceanographic data useful on the spot.	This may be of overriding importance in considering military objectives.	A moth zig zagged along the path through otto's garden.
Buying a thoroughbred horse requires intuition and expertise.	His technique is ample and his musical ideas are projected beautifully.	Al received a joint appointment in the biology and the engineering departments.
The high security prison was surrounded by barbed wire.	Count the number of teaspoons of soysauce that you add.	While waiting for chipper she crisscrossed the square many times.
No girl would go this far to fool a man so she could kill him.	Northern liberals are the chief supporters of civil rights and of integration.	That stinging vapor was caused by chloride vaporization.
Whether historically a fact or not the legend has a certain symbolic value.	The library has open shelves even in the unbound periodical stockroom.	Steph could barely handle the psychological trauma.
In every major cloverleaf traffic sometimes gets backed up.	Davy mathews it's disgusting the way you're always eating.	As a precaution the outlaws bought gunpowder for their stronghold.
The entire length of the street could be raked with rifle fire from this barn.	Severe myopia contributed to ron's inferiority complex.	May i order a strawberry sundae after i eat dinner.
And never show my face or my truck around here again.	Any organism that falters or misperceives the signals or weakens is done.	Valley lodge yearly celebrates the first calf born.
Then again there's always that lovely old pastime of hooking or braiding rugs.	A young mouse scampered across the field and disappeared.	What possessed you to tell me a clotheshorse would be a good idea.
Employee layoffs coincided with the company's reorganization.	The antithesis of the ecumenical and the local then no longer exists.	The social and psychological consequences of this continue to affect the area.
Make a paste of brown sugar and mustard and spread lightly over scored surface.	In many of his poems death comes by train a strongly evocative visual image.	In most cases we recognize certain words persons animals or objects.
Cooperation along with understanding alleviate dispute.	Gregory and tom chose to watch cartoons in the afternoon.	Whoever cooperates in finding nan's cameo will be rewarded.
Such legislation was clarified and extended from time to time thereafter.	But none of this could soothe the exacerbated nerves.	He waited until they were inside the elevator and then said now what do we do.
Coffee is grown on steep jungle like slopes in temperate zones.	We like bleu cheese but victor prefers swiss cheese.	My desires are simple give me one informative paragraph on the subject.
Alice i tell you my feet are like chopped beefsteak.	Almost everybody in the senior class is married students say dogmatically.	They all agree that the essay is barely intelligible.
It moved in a silver arc toward his throat then veered downward.	His legs pumped furiously his long black hair streamed out behind him.	If we left one we'd have to wipe it for fingerprints.

The rich should invest in black zircons instead of stylish shoes.	And the surface is driven back in its very surfaceness only by this contrast.	Since a fall or blow might have caused it a cold pack was usually first aid.
Positive results start when it goes towards the hand you use to make your mark.	The preschooler couldn't verbalize her feelings about the emergency conditions.	Are we as safe as we should be from such a disaster.
Before deriving this formula we explain what we mean by problems of this kind.	My sincere wish is that he continues to add to this record he sets here today.	Neither his appetites his exacerbations nor his despair were kin to yours.
The giant redwoods shimmered in the glistening sun.	The cow wandered from the farmland and became lost.	We have become amateur insurance experts and fine feathered yard birds.
Privately he created and magnified an image of himself as a hired assassin.	Both loved the out of doors including mountain climbing and horseback riding.	And their chroniclers are not the dramatic poets but the prose novelists.
According to my interpretation of the problem two lines must be perpendicular.	In the course of its inquiry it took testimony from only seven witnesses.	Seamstresses attach zippers with a thimble needle and thread.
Although they drew light ground fire they saw no signs of activity.	Nonprofit organizations have frequent fund raisers.	Boys and men go along the riverbank or to the alcoves in the top arcade.
The response of reaction is dominated by a concern for what is vanishing.	The fat man has trouble buying life insurance or has to pay higher premiums.	On all sides doors were being slammed in his face.
Which long article was opaque and needed clarification.	To what extent such low density applies to micrometeorites is unknown.	The patient and the surgeon are both recuperating from the lengthy operation.
The haunted house was a hit due to outstanding audio visual effects.	Those answers will be straightforward if you think them through carefully first.	This was chiefly because of the bluish white autofluorescence from the cells.
Biologists use radioactive isotopes to study microorganisms.	At the left is a pair of dressy straw pumps in a light but crisp texture.	He reached out and felt the bath towel hanging on the towel rack over the tub.
Thus far the advances made have been almost entirely along functional lines.	There was a gigantic wasp next to irving's big top hat.	With her sharp tongue she'd have cut his pomposness to ribbons.
However when labor disputes arise its provisions come clearly into play.	Confusion became chaos each succeeding day brought new acts of violence.	His tough honesty condemns him to a solitary and difficult existence.
Later we shall see what happened when an emperor took this idea too literally.	Thirty five they rode at a measured pace through the valley.	Technical writers can abbreviate in bibliographies.
A clearly recognized exception is a statutory merger or consolidation.	Hiding out like this won't get him anything except more trouble or bullet.	I gave them several choices and let them set the priorities.
The water contained too much chlorine and stung his eyes.	An adult male baboon's teeth are not suitable for eating shellfish.	Begin the examination of a site with a good map and aerial photos if possible.
These exclusive documents must be locked up at all times.	Butterscotch fudge goes well with vanilla ice cream.	Index words and electronic switches may be reserved in the following ways.
Thus there is a clearer division of authority administrative and legislative.	This is a significant advance but its import should not be exaggerated.	Todd placed top priority on getting his bike fixed.
He was above all a friend seeker almost pathetic in his eagerness to be liked.	So we note approvingly a fresh sample of unanimity.	No she would not pretend modesty but neither must she be crudely bold.
Roy ignored the spurious data points in drawing the graph.	Whosoever violates our rooftop the legend states can expect maximal sorrow.	A sailboat may have a bone in her teeth one minute and lie becalmed the next.
The new suburbanites worked hard on refurbishing their older home.	He played basketball there while working toward a law degree.	He had accordingly cultivated eccentricity to the point of second nature.
The frightened child was gently subdued by his big brother.	No manufacturer has taken the initiative in pointing out the cost involved.	The desire and ability to read are important aspects of our cultural life.
A complete plan we have made limited application of the parallel ladder plan.	The public is now armed with sophistication and numerous competing media.	The barracuda recoiled from the serpent's poisonous fangs.
He liked to nip ear lobes of unsuspecting visitors with his needle sharp teeth.	Two miles northeast then five miles southwest that sort of thing.	He picked up nine pairs of socks for each brother.
Brush fires are common in the dry underbrush of nevada.	The dead spirits occupied a prominent place in every hope and in every fear.	Behind him billowed a small pungent cloud of smoke.
Suburban housewives often suffer from the gab habit.	Will you please confirm government policy regarding waste removal.	Perhaps this is what gives the aborigine his odd air of dignity.
Thereupon followed a demonstration that tyranny knows no ideological confines.	Research into several cultures has proven her position to be a mistaken one.	She had jumped away from his shy touch like a cat confronted by a sidewinder.
Faces may be made into candles by filling with melted wax and a wick.	In most discussions of this phenomenon the figures are substantially inflated.	C'mon he whispered four levels about three feet down so don't fall.
The sudden solitude had lost its momentary charm and become oppressive.	Vital questions would be quickly answered according to a preprepared agenda.	The mayan neoclassic scholar disappeared while surveying ancient ruins.
They also furnish proof that in modern war message sending must be monitored.	But problems cling to pools as any pool owner knows.	Personal predispositions tend to blunt the ear and in turn the voice as well.
Only rarely is attention given to accurate progress reports and evaluation.	Both have excellent integration of their fiscal tax collection year calendars.	Need for novelty may be a symptom of cultural fatigue and instability.
The mixing head moves back and forth slowly across the width of the receptacle.	First they wanted to clarify a tantalizing bizarre enigma.	Shell shock caused by shrapnel is sometimes cured through group therapy.

The groundhog clearly saw his shadow but stayed out only a moment.	Two cars came over a crest their chrome and glass flashing.	He really crucified him he nailed it for a yard loss.
The trouble is that like many symbols it doesn't seem a very realistic one.	He ripped down the cellophane carefully and laid three dogs on the tin foil.	The flat bottomed boat swung slowly to the pull of the current.
Scientific progress comes from the development of new techniques.	The cigarettes in the clay ashtray overflowed onto the oak table.	The most recent geological survey found seismic activity.
Men believed they could control nature by obeying a moral code.	He spoke briefly sensibly to the point and without oratorical flourishes.	I'll have a scoop of that exotic purple and turquoise sherbet.
But it must be remembered that the plan should not be oriented geographically.	Come on let's hurry down before they lock up for the day.	The boston ballet overcame their funding shortage.
Yet the spirit which lives in community is not identical with the community.	Wine glass heels are to be found in both high and semi heights.	Naturally no woman can ever completely monopolize the sexual initiative.
The local drugstore was charged with illegally dispensing tranquilizers.	Flying standby can be practical if you want to save money.	Assume for example a situation where a farm has a packing shed and fields.
It made no difference that most evidence points to an opposite conclusion.	Rob made hungarian goulash for dinner and gooseberry pie for dessert.	Shall we flip a coin to see which of us goes first.
Jokes cartoons and cynics to the contrary mothers in law make good friends.	Be careful that you keep adequate coverage but look for places to save money.	But fenced or unfenced no pool side is the place for running or horseplay.
Another field had given him fame enough to satisfy any egotist.	The cowboy's humorous name for a cow givin' milk was a milk pitcher.	For wasp stings onion juice obtained by scraping an onion gave quick relief.
He may have a point in urging that decadent themes be given fewer prizes.	Now don't tell me what a good ball player you are.	They know little about their machinery beyond mechanical details.
The legislature met to judge the state of public education.	He is forced to play for little money and must often take another job to live.	Some have walked through pain and sorrow to bring you their message of hope.
The thick elm forest was nearly overwhelmed by dutch elm disease.	Needless to say my art suffered drastically during this turbulent period.	Once you finish greasing your chain be sure to wash thoroughly.
When peeling an orange it is hard not to spray juice.	They stayed at hotels and boardinghouses or at private homes.	Internal national responsibility now a truism need not be documented.
Ironically enough in this instance such personal virtues were a luxury.	Far more frequently overeating is a result of a psychological compulsion.	Solid concrete blocks relatively heavy and dense are used for this shelter.
The overweight charmer could slip poison into anyone's tea.	Along the main thoroughfares hardly a house had not been peppered.	We experience distress and frustration obtaining our degrees.
The lack of heat compounded the tenant's grievances.	Selecting bunks by economic comparison is usually an individual problem.	Yeah seems so don't it the boy laughed hugging her close.
He showed puny men attacked by splendidly tyrannical machines.	He can for example present significant university wide issues to the senate.	The moisture in my eyes is from eyedrops not from tears.
Your voice is delightful he approved with a warm smile.	Differences were related to social economic and educational backgrounds.	He crossed the next meadow and climbed a tree where the jungle trail resumed.
This truth that the moral law is natural has other important corollaries.	If they are not ellipsoids the conclusions will be a reasonable approximation.	But he was very much like his associates in his hatred of camp routine.
His blue eyes sought the shimmering sea of haze ahead.	The single curve line represents a specific formulation in a test example.	He saw a pint sized man with a graying spade beard and an unusually large head.
The word means it won't boil away easily nothing else.	These air or gas bubbles make highly functional thermal barriers.	The proof that you are seeking is not available in books.
A screwdriver is made from vodka and orange juice.	Unit prices for state vehicles are invariably lower than to the general public.	Every single problem touched on thus far is related to good marketing planning.
While one element is announcing progress another is delineating its problems.	A covered container such as a kitchen garbage pail might do as a toilet.	Why the hell didn't you come out when you saw them gang up on me.
Many wealthy tycoons splurged and bought both a yacht and a schooner.	Chip postponed alimony payments until the latest possible date.	A concept of responsibility is in process of articulation and establishment.
A domestic automatic washer that will give equivalent results may be used.	The compounds are divided according to composition into seven categories.	Draw every outer line first then fill in the interior.
Usually they titter loudly after they have passed by.	Chocolate and roses never fail as a romantic gift.	It was as blissful and fulfilling a night as any bride ever experienced.
Yes indeed we too can see a warlike host of infidels encamped against us.	Lifting her skirts she climbed in never relinquishing her grip on his arm.	Again these blocks were set in resin saturated glass cloth and nailed.
Often they are able to get in only because the area is declining economically.	His prescription hot and cold compresses to increase her absorption of water.	Some make beautiful chairs cabinets chests doll houses etc.
For girls the overprotection is far more pervasive.	When mold has more than one design cavity make individual paper patterns.	One upmanship is practiced by both sides in a total war.
The shot reverberated in diminishing whiplashes of sound.	Some observers speculated that this might be his revenge on his home town.	The tragic stage is a platform extending precariously between heaven and hell.
Bruises and black eyes were relieved by application of raw beefsteak.	She knew she was feeling afraid and inwardly laughed at herself.	However this inaugural feast did its sponsors no good whatever.

Superior new material for orthodontic work is another result of research.	He strolled back to the door whistling softly hands still clasped behind him.	Her eyes were glazed as if she didn't hear or even see him.
She found herself able to sing any role and any song which struck her fancy.	In most cases these soils are taken up as liquids through capillary action.	That doctrine has been accepted by many but has it produced good results.
A smile pulled at the lower strip of adhesive tape.	The orchestra was obviously on its mettle and it played most responsively.	Then he would realize they were really things that only he himself could think.
Space charge influences will also decrease at increased voltages.	She served for a number of years without pay beyond her travel and maintenance.	His superiors had also preached this saying it was the way for eternal honor.
In tradition and in poetry the marriage bed is a place of unity and harmony.	He injected more vitality into the score than it has revealed in many years.	Somehow we old timers never figured we would ever retire.
Sprouted grains and seeds are used in salads and dishes such as chop suey.	Its sphere is that of royal courts dynastic quarrels and vaulting ambitions.	Not without good reason has the anatomical been called jocular journealese.
Very peculiar retribution indeed seems to overtake such jokers.	His portrayal of an edgy head in the clouds artist is virtually flawless.	The theory the idea behind our design is modular units or panelization.
You think somebody is going to stand up in the audience and make guilty faces.	His name became synonymous with cold blooded cruelty.	One of the most common of camp maladies was diarrhoea.
No group of officers came in for more spirited denunciation than the doctors.	We seemed to be witnessing the population explosion right in our own backyards.	Being based on so few events these results are of dubious validity.
For me it has more of both elements than the majority of its competitors.	Adults take a long time to convince and you are thwarted if you try to push.	You could also say that in these pamphlets is a relieving quality of maturity.
But considered within technical astronomy a different pattern can be traced.	Neither are beds thanks to air mattresses and sleeping bags.	Rabies were cured or prevented by madstones which the pioneer wore or carried.
Like a fair number of bootleggers he disliked alcohol.	How much and how many profits could a majority take out of the losses of a few.	And he was handsome despite the long thin scar that slanted across his cheek.
She encouraged her children to make their own halloween costumes.	His first glimpse of the ranch house across the brushy swells told him nothing.	They were shown how to advance against an enemy outpost atop a cleared ridge.
Have a test run on the family first to be sure timing and seasoning are right.	As coauthors we presented our new book to the haughty audience.	To keep 'em scattered somewhat and yet herd 'em was called loose herdin'.
We have also seen the power of faith at work among us.	Last year's gas shortage caused steep price increases.	To use these new ways in daily life is the last step.
We now generalize these ideas for general binomial experiments.	An area sheltered from strong winds may be highly desirable for recreation use.	We would lose our export markets and deny ourselves the imports we need.
A cardboard pattern cut to fit inside holder will help to prevent warping.	These were thought to represent regenerating fibers.	The wrinkled mouth laughed revealing astonishingly strong white teeth.
It seems that open season upon veterans' hospitalization is once more upon us.	Why don't they tell me themselves if it bothers them.	Cory attacked the project with extra determination.
Now you know she could've but she isn't that kind of girl.	It had assumed the terrifying inertia of inanimate matter.	It would be well to show the populace how we deal with adulterers.
In an ideological argument the participants tend to thump the table.	The narrow fringe of sadness that ran around it only emphasized the pleasure.	Worse his present crew included five men who had sailed with him before.
Her face seemed to float in an implausibly bright shaft of sunlight.	Sometimes soldiers wrote letters while bullets were whizzing about their heads.	The thought came back the one nagging at him these past four days.
Manual leveling requires an appropriate display of the accelerometer outputs.	The other patrons were taxi drivers and art students and small shopkeepers.	State numbering laws differ from each other in many ways.
But the information on the dynamics of population was often quite misleading.	We will achieve a more vivid sense of what it is by realizing what it is not.	Originals are not necessarily good and adaptations are not necessarily bad.
The odor here was more powerful than that which surrounded the town aborigines.	She drank greedily and murmured thank you as he lowered her head.	In my place you'd follow such advice as you give me.
I took her word for it but is she really going with you.	Sometimes although by no means always these are indeed alkaline.	Mike was of legal age and presumed able to defend himself in the clinches.
As we observe moral law and physical law they appear as being inevitable.	He was busy he said in having someone submit to a monkey gland operation.	Well i guess i ought to dust out that desk anyhow.
Our hypothetical other bum who killed him would have turned out his pockets.	There should be no reason to misinterpret or ignore the intent of this letter.	No amount of ballyhoo will cover up the sordid facts.
We may say of some unfortunates that they were never young.	The word also made him feel hate sincere hate for those so labeled.	Impressions often appear in a symbolic form and cannot be taken at face value.
There are several sources of evidence on the micrometeorite environment.	Ice baths electric shocks lashings wild dogs testicle crushers.	Caution continuous administration is not recommended for lactating cows.
Ideally he knew it should be preceded by concrete progress at lower levels.	The decking is quarter inch mahogany marine plywood.	The battery median grade equivalent was used in data analysis in this study.
There are optimal humidity requirements for various agents when airborne.	It was the story of the rhinoceros fight all over again.	Soil redeposition is evaluated by washing clean swatches with the dirty ones.
What a discussion can ensue when the title of this type of song is in question.	Meanwhile fishermen took advantage of them to pull up whoppers.	The season between spring and summer belongs to life in its carefree aspect.

Opaque cantaloupe and transparent wood brown were used.	Wooded stream valleys in the folds of earth would be saved.	He didn't tell her the truth he now freely admitted to himself.
Now the problem is presented piecemeal and sometimes contradictorily.	Questions came to me from all sides about my world citizenship activities.	He did not however settle back into acquiescence with things as they were.
For roast insert meat thermometer diagonally so it does not rest on bone.	Again the analyticity of the two curves guarantees that such intervals exist.	Lips pursed mournfully he stared down at its crazily sagging left side.
During one reading an image appeared of a prisoner in irons.	Is there any word you would like to offer in your own defense.	Not very well behaved is she to run out on a play mate.
We did not accept the diagnosis at once but gradually we are coming to.	Another brand of indefinite reference arises out of the use of the double verb.	The failure to keep these two usages distinct presents hazards to the reader.
If you destroy confidence in banks you do something to the economy he said.	New self deceiving rags are hurriedly tossed on the too naked bones.	Somewhere birds were sweetly calling were answered.
This process is especially difficult since gyro drifting is typically random.	Keep your seats boys i just want to put some finishing touches on this thing.	The world is constantly changing what was new yesterday is obsolescent today.
Although my shot killed his horse he rolled off the bale on top of me.	It latches when you close it so stay as long as you like.	Other interpretations present the music as an essentially intimate creation.
She had no way of knowing in advance whether an opportunity for murder existed.	No more startling contrast to a system of sullen satellites could be imagined.	They make gin saws and deal in parts supplies and some used gin machinery.
A chaw of tobacco put on an open wound was both antiseptic and healing.	No antigen was detectable in certain dark spherical areas in most cells.	It was like finally getting into one's own nightmares to punish one's dreams.
One species of ambiguity tries to baffle by interweaving repetition.	Though brief it has a sharp dramatic edge and great poignancy.	The continuing modernization of these forces is a costly but necessary process.
Ran away on a black night with a lawful wedded man.	If you use parking attendants can they be replaced by automatic parking gates.	A monstrous shadow fell across the illuminated wall distorted and indefinable.
It will accommodate firing rates as low as half a gallon an hour.	One of these is the solidarity and the confidential relationship of marriage.	Other morphological physical and optical property values are also given.
She took it grudgingly her dark eyes baleful as they met his.	The highest rated non supervisory engineering title is research engineer.	What had been the ambassador's suite was now jagged walls of blackened brick.
He is not talking in the main about probabilities risks and danger in general.	A kerosene shampoo seems a heroic treatment but it did the job.	They should live in modest circumstances avoiding all conspicuous consumption.
Measured performance characteristics for this experimental tube will be listed.	One of the most desirable features for a park are beautiful views or scenery.	The smell is sexual but so powerfully so that a civilized nose must deny it.
Death reminds man of his sin but it reminds him also of his transience.	He merely said any good decorator these days can make you a tasteful home.	Thinking the evidence insufficient to get a conviction he later released him.
Conservatism and traditionalism seem implied by what has just been said.	Thirty five military and civilian students received laboratory training.	They did not know who they were or know their own worth.
Pretty girls among them with blonde hair and pert faces.	He stared at the far morning expecting a pendulum to swing across the horizon.	Also make sure thermometer does not touch the revolving spit or hit the coals.
If it ever got behind me the beep turned to a buzz.	We flew in rickety planes so overloaded that we wondered why they didn't crash.	Tetanus could be avoided by pouring warm turpentine over a wound.
So if anybody solicits by phone make sure you mail the dough to the above.	But they would reconsider it they assured him if he would rewrite it.	Clapping spurs to the bronc he set off at a sharp canter with growing alarm.
She took it with her wherever she went she chose it.	The old woman arose stiffly and led me to a clearing where a small hut stood.	There was no confirmation of such massive assaults from independent sources.
The walls bulged the floor trembled the windowpanes rattled.	Sewing brings numbness writes what makes my hands numb when sewing.	For sweet sour sauce cook onion in oil until soft.
Running around in the moonlight almost naked and slugging a man with a rock.	He remembered the last time he had eaten actual eggs from an actual pan.	The simplest kind of separate system uses a single self contained unit.
You could burn down this whole mountainside with a fire that size.	There are many such competently anonymous performances among the earlier poems.	If a concessionaire runs the cafeteria keep an eye out for quality and price.
Ralph prepared red snapper with fresh lemon sauce for dinner.	This staff deserves a lot of credit working down here under real obstacles.	Tiny bodies dropped onto a dry leaf made a pile as big as a small apple.
No chemical fertilizers and poisonous insecticides and fungicides are used.	The total of these three volumes is the final combustion chamber volume.	We were off the road gleaming barbed wire pulling taut.
Yet it exists and has an objective reality which can be experienced and known.	They inhabit a secret world centered on go codes and gold phones.	Latest models serve hot meals at reasonable prices and at a profit to you.
Both eventualities are possible logically but practically they are impossible.	The system may break down soon so save your files frequently.	But why is it necessary to reproduce the retinal image within the brain.
He daydreamed on the rock while she swam and splashed around.	Bathing the itching parts with kerosene gave relief and also killed the pests.	There is little doubt that the students benefit from vocational education.
The gunman nodded slipping the picture into his breast pocket saying nothing.	They were not yet prepared to accept it as irremediable.	With any luck at all he could easily find a flowerpot.
Evidence that other sources of financing are unavailable must be provided.	We send shovels cement nails and corrugated iron for roofs.	The avocado should have a give to it as you hold it when it is ripe.

Prompted by a guilty urge he had disobeyed the order of a man he respected.	He will say that our country is even now a homogeneous community.	This girl soon drops the bourgeoisie psychiatrist who disapproves of her life.
He slipped outside hugging the walls of buildings and dodging into doorways.	And men also used vacuum cleaners in both rooms sucking dust up once more.	The dimensions of these waves dwarf all our usual standards of measurement.
Promptly at seven he would clatter out of the court with twelve in the tallyho.	Now a distinguished old man called on nine divinities to come and join us.	He hoped they would put in somewhere way way down in the earth.
The population can thereby replenish itself and actually grow larger.	They offered no opinions volunteered nothing betrayed no emotions.	These needs usually concern the reduction of guilt and some relief of tension.
Microorganisms are often responsible for the rapid spoilage of foods.	One of the problems associated with the expressway stems from the basic idea.	Only incomplete imperfect things move towards what they lack.
We'll both be blowing town tomorrow so we won't be moving in on you.	It required an energy he no longer possessed to be satirical about his father.	No question ruffles him or causes him to hesitate.
Flaxseed poultices and mustard plasters still are used by some persons.	All that time rifle barrels were pointing unwaveringly at his head and body.	Steam baths writes do steam baths have any health value.
In time she presents her aristocratic husband with a coal black child.	The straight line would symbolize its uniqueness the circle of universality.	Even then if she took one step forward he could catch her.
Would have been easy to identify as opium by its odor.	This is a problem that goes considerably beyond questions of salary and tenure.	Quince seed gum is the main ingredient in wave setting lotions.
Half slyly he enjoyed seeing her stoop to lift the things.	We hope that he will execute it in a manner that will entitle him to credit.	A site may also be attractive just through the beauty of its trees and shrubs.
Her face turned pink with pleasure and a smothered cough.	Would a camera club be useful in taking pictures pertinent to plant safety.	Family loyalties and cooperative work have been unbroken for generations.
Would a blue feather in a man's hat make him happy all day.	He reached once more into the carpet bag and brought up a package of wieners.	There was a grunt curiously inarticulate like that of an animal in pain.
Insulate weatherstrip double glaze to the maximum.	He wanted to show the town what happened to anyone who tried to start trouble.	Ants carry the seeds so better be sure that there are no ant hills nearby.
You should firmly insist that no bobby pins or hair pins be worn in the water.	In some measure they depend upon the structure of individual personality.	We congratulate the entire membership on its record of good legislation.
On unoccupied roadway the bottle shattered into a small amber flash.	The name fell with lazy affectionate remembrance from her lips.	When he finally did he had to duck his head quickly away as the pitch came in.
Afraid you'll lose your job if you don't keep your mouth shut.	But to the infuriation of scientists for no known reason not all of them did.	Meanwhile the enemy will capitalize on our fears if he can.
Further it has its work cut out stopping anarchy where it is now garrisoned.	It has multiple implications and possible headaches for your marketing program.	One more muddleheaded play like that one and they'd be leading him away.
Traffic frequently has failed to measure up to engineers' rosy estimates.	Ambiguity arises when the pronoun it carries a twofold reference.	However the aircraft which we have today are tied to large soft airfields.
We know that actors can learn to portray a wide variety of character roles.	The ward was a small one four beds kept reserved for female alcoholics.	The sculptor looked at him bug-eyed and amazed angry.
But to continue to divorce advanced students from reality is inexcusable.	But this doesn't detract from its merit as an interesting if not great film.	Then may I ask where these muddy foot prints came from.
In the lighted interior he saw other men and women struggling into their wraps.	Her debut over perhaps the earlier scenes will emerge equally fine.	He straightened up alert now as the buffalo hunter came closer.
The crisis later on when debts seemed about to overwhelm me.	Something else distracted him yet there was no sound only tomblike silence.	Let the orthodontist decide the proper time to start treatment he urges.
Forty seven states assign or provide vehicles for employees on state business.	This he added brought about petty jealousies and petty personal grievances.	It is possible to make a few generalizations about the six giants themselves.
This has been attributed to helium film flow in the vapor pressure thermometer.	As his feet slowed he felt ashamed of the panic and resolved to make a stand.	Perhaps one bored holes in the stone with some kind of an electric gadget.
The keelson made of two three inch widths is next installed.	Cleaned cloth must be protected against the redeposition of dispersed soil.	Brief snips of actual events were shown parades dances street scenes.
His black hat with its wide brim high crown and fur trim rode high.	Get copper or earthenware mugs that keep beer chilled or soup hot.	Two clotted balls the color of mucus rolled between fiery lids.
Pretend ham make criss cross gashes on one side of skinless frankfurters.	The platform accelerometers must be slightly modified for this procedure.	They are not true because scientists or prophets say they are true.
The batting average of one success out of seven increased to one out of three.	And his relatively small hands and feet gave him an almost delicate appearance.	But from the start they had two important ingredients sincerity and realism.
From it spokes of order and degree led to the outward rim of the common man.	Resolved that the anti slavery sentiment is becoming ripe for resolute action.	But if she wasn't interested she'd just go back to the same life she'd left.
He looked over at him lying there asleep and he felt a wave of revulsion.	In the winter hibachi in the kitchen or grill over the logs of the fireplace.	Higher toll rates also are helping boost revenues.
He would offer no theory to account for her murder.	He drove sensual patterns off carefully shaving his long upper lip.	More often these offices are restricted to the gathering of empirical data.
In these damp circumstances he was an odds on bet to develop pneumonia.	The problem of solidarity and morale again involves the concept of values.	Maybe you and me will girlie but these two ain't goin' nowhere.

Was it a hysterical release from the long strain of vigilance of those weeks.	Both the conditions and the complicity are documented in considerable detail.	Of particular importance is the study of the actions of drugs in this respect.
Advantages a farm provides a wholesome and healthful environment for children.	His eyes were dark fluid fearful and he gave a sigh as my knife went in.	It does not indicate loose management ineffective controls or poor policy.
These curves were derived by an analysis of extensive skywave measurement data.	Don't they still call you junior as though you're about ten years old.	Fall slowly forward onto the hands and let the body down to rest on the floor.
Except for those minutes in her room he had lost touch with her as a reality.	Replace it with the statue of one or another of the world's famous dictators.	The need for reupholstering redecorating repainting becomes more infrequent.
The beatniks crave a sexual experience in which their whole being participates.	So somebody else knew what would happen to her father's money if she died.	If the other pilots were worried they did not show it.
Our first necessity at the very outset of war is post attack reconnaissance.	To some extent predispositions are shaped by exposure to group environments.	In earlier years the preservation of food was essentially related to survival.
Was she just naturally sloppy about everything but her physical appearance.	Sometimes strong stress serves to focus an important secondary relationship.	Hired hard lackeys of the warmongering capitalists.
He holds that goodness and badness lie in feelings of approval or disapproval.	He murmured to himself with firmness no surrender.	We scour literature for them here we find stored the wisdom of great minds.
And these hardy travelers are not unappreciated today.	The revolution now under way in materials handling makes this much easier.	With those paintings of big constructions crashing down he felt he could stop.
He didn't figure her at all and if he found out a woman it'd be bad.	It offered to surrender its right to exclusive trade but asked an indemnity.	He may not rise to the heights but he can get by and eventually be retired.
Her position covers a number of daily tasks common to any social director.	A flash illumined the trees as a crooked bolt twiggid in several directions.	Biological warfare is considered to be primarily a strategic weapon.
They moved toward the skiffs with shocking eagerness elbowing and shoving.	Radio reception is cut down by the shielding necessary to keep out radiation.	Then he fled not waiting to see if she minded him or took notice of his cry.
Avowed atheists or freethinkers are so rare as to be a curiosity.	This big flexible voice with uncommon range has been superbly disciplined.	She greeted her husband's colleagues with smiling politeness offering nothing.
Come sit he repeated motioning to the piled hay bags over the pig leavings.	Before that we lumber dealers were working almost single handed on the problem.	She smiled and teeth gleamed in her beautifully modeled olive face.
Meanwhile three great terrible forces were coagulating and crystallizing.	Residential associations struggle to insulate themselves against intrusions.	You'd think her stomach would've got used to it in three weeks.
Not good looking but self confident and wise so that it made her attractive.	Roleplaying used for analysis follows these general steps leading to training.	To prepare mustard cream blend mustard with enough water to make a thin paste.
After another long pause he asked how many people know who they are.	Within a system however the autonomy of each member library is preserved.	The old shop adage still holds a good mechanic is usually a bad boss.
Beer generally fermented from barley is an old alcoholic beverage.	When you're less fatigued things just naturally look brighter.	Quite often honeybees form a majority on the willow catkins.
They were both walking towards each other unhurried.	He must rearrange matters so that two performers do not bump into each other.	Wet also were the marine's fatigues and the face had an oily film.
They weren't as well paid as they should have been.	It could take place tomorrow night or it might occur months from now.	When we left washington his son tad was ill and mrs lincoln hysterical.
The long and ever increasing column of sportsmen is now moving into a new era.	The bacteria formed typical activated sludge floc.	The filtered air benefits allergies asthma sinus hay fever.
Are planning and strategy development emphasized sufficiently in your company.	His talk turns to what he calls the mess or sometimes this buzzing confusion.	Conceivably the submarine defense problem can be solved by sufficient forces.
Two women who had been chattering like parrots were struck dumb.	In news items a man is less often shot in the body or head than in the suburbs.	This is going to be a language lesson and you can master it in a few minutes.
He doesn't want her to look frowningly at him or speak to him angrily.	But briefly the topping configuration must be examined for its inferences.	A vigorous program existed in skiing skating sports and overnight hiking.
This is the second year your mother's donated my fishing tackle to the bazaar.	When did women begin to assert themselves sexually.	The good man chortled appreciatively and decided the trip was worth his time.
This explains some group ends and provides a justification of their primacy.	Computers are being used to keep branch inventories at more workable levels.	Large injection molded letters are also available for sign installations.
The bristles are soft enough to massage the gums and not scratch the enamel.	There are canoes ideal for fishing in protected waters or for camping trips.	They'll move around that rock all day following the shade.
It takes a great deal of sophisticated thought to get the impact of this fact.	His problem concerns longitudes latitudes and angular velocities.	Did you know he is advertising his ham radio equipment for sale this weekend.
The wheel of social life spun around the royal or aristocratic centre.	But this esoteric doctrine was lost in the shuffle to acquire special powers.	We do not arrive at spatial images by means of the sense of touch by itself.
There are people who travel long distances to assure my continued existence.	The same shelter could be built into an embankment or below ground level.	Everything in the final analysis reduced itself to sexual symbolism.
Is a relaxed home atmosphere enough to help her outgrow these traits.	The elementary school child grows gradually in his ability to work in groups.	We're lost and burning up already she bit out tensely.
Probably around midnight give or take an hour either way.	She saw me and sat down beside me three feet away.	But i'm so sunburned that every move i make is agony.

Closer still regular barricades of barbed wire hung on timber supports.	Cattle which died from them winter storms were referred to as the winter kill.	Hiring the wife for one's company may win her tax aided retirement income.
Paperweight may be personalized on back while clay is leather hard.	Blowers should be operated periodically on a regular schedule.	There is definitely some ligament damage in his knee.
Another stock vaudeville gag ran mother is home sick in bed with the doctor.	He enlisted a staff of loyal experts and of many zealous volunteers.	We often say of a person that he looks young for his age or old for his age.
It cost us a hundred thousand dollars and thirty days lost time to fix them.	Or borrow some money from someone and go home by bus.	Almonds and pistachio nuts are not so high in oil but are rich in protein.
Castor oil made from castorbeans has gone out of style as a medicine.	He went on to personal bequests a list of names largely unknown to him.	This is no assignment for a frivolous girl she assures him.
The stepmother almost without exception has been presented as a cruel ogress.	You're not living up to your own principles she told my discouraged people.	He'd had no idea how unhappy his sweet peach had been.
Women didn't use white face powder nowadays he recalled.	I always say you've got a wonderful husband miss margaret.	It is one of the rare public ventures here on which nearly everyone is agreed.
The larvae kept warm by the queen are full grown in about ten days.	She always did before and showed the utmost confidence in whatever we advised.	As he had longed to be he became the echo of a saga.
This viscosity of the material in the drops is of course not negligible.	High so it only bounce harmlessly but loudly off a car's steel roof.	Maybe it's taking longer to get things squared away than the bankers expected.
Can your insurance company aid you in reducing administrative costs.	One can only speak of what is in front of him and that now is simply the mess.	Push ups push ups are essential but few have the strength for them at first.
In simpler terms it amounts to pointing the platform in the proper direction.	From this motor pool personnel develop other meaningful and related data.	This one came a bit high at thirty thousand or more.
Most of our aid will go to those nearing self sufficiency.	The shock therapies act likewise on the hypothalamic balance.	The knifelike pain in his groin nearly brought him down again.
To be passive to be girlishly shy was palpably absurd.	Look for these features which may mean you can save duplicate coverage.	When he left she knew she would never see him again.
Under this law annual grants are given to systems in substantial amounts.	But our stumping tour of the south wasn't all misery.	My beloved ward my perennial gadfly said the whining voice.
His sinuous melody is a sort of naive transcendence of all experience.	Anybody carrying anything that might hide a rifle.	Woe betide the interviewee if he answered vaguely.
Program note reads as follows take hands this urgent visage beckons us.	In either case they do not appreciate the private detective's zeal.	X ray films of the vertebral column showed progression of the demineralization.
Microscopically there was emphysema fibrosis and vascular congestion.	In a new house generous roof overhangs are a logical and effective solution.	Are you utilizing vending machine proceeds to help pay for your program.
This is what necessitates the nonsystematic character of his astronomy.	Individual human strength is needed to pit against an inhuman condition.	Knows the score with a supreme effort he broke it off.
Why couldn't they have dumped him off on someone else.	His successors have adopted the opposite alternative.	Crooked overlapping twisted or widely spaced teeth.
His artistic accomplishments guaranteed him entry into any social gathering.	They played crack the whip a few minutes without mishap.	

Table 13. 728 randomized scripted prompts for version 1.0 and 2.0, from which each participant only responded to 10.

R. Video Identity Matching

Table 14. Video identity pairs found through the minimal cosine distance between mean CLIP embeddings of five video frames.

identity_1	identity_2	identity_1	identity_2	identity_1	identity_2
0	1	85	86	143	144
2	3	87	192	146	171
4	5	88	126	147	154
6	7	89	106	148	155
8	9	90	194	150	151
10	11	91	119	153	163
12	13	92	184	156	157
14	15	93	207	158	159
16	17	94	191	160	161
18	19	95	201	166	167
20	21	96	142	168	178
22	23	97	98	169	202
24	25	99	100	175	198
26	27	101	174	176	197
28	29	102	196	179	218
30	31	103	134	180	181
32	33	104	193	182	183
34	35	105	177	186	215
37	38	107	170	187	217
39	40	108	109	188	211
41	42	110	111	189	190
43	44	112	113	195	221
45	46	114	115	199	200
47	48	116	124	203	204
49	50	117	208	209	210
51	52	118	165	212	213
53	54	120	129	222	258
55	56	121	214	223	419
57	58	122	137	224	230
59	60	123	162	225	411
61	62	125	220	226	270
63	64	127	172	227	400
65	66	128	164	228	480
67	68	130	152	229	253
69	70	131	149	231	414
71	72	132	206	232	361
73	74	133	173	233	462
75	76	135	205	234	264
77	78	136	219	235	393
79	80	138	145	236	467
81	82	139	185	237	383
83	84	140	141		