SYNTHETIC MEDIA DETECTION · AUDIO AND VIDEO

DISCERN-AI

The signature of a living body.

Benchmark results. Current edition, 9 November 2026.

Paul Samuel Guarino

Inventor · East Northport, New York · 41.176hz@gmail.com

THE PRINCIPLE

A forgery can copy how a person looks. It cannot copy how a living body holds itself together.

A living voice is a closed loop: the speaker hears themselves and corrects, continuously. A living face is held together by coupled rhythms that move as one system. Generated media is rendered. Its variation is produced, not regulated.

Most detectors learn the fingerprints of specific generators, and those fingerprints change with every new release. DISCERN-AI measures the source instead: the relationship that only a living system has.

VIDEO · AI-GENERATED FACES

Real and generated video, separated at AUC 0.974.

0.974
AUC, all clips
1,312 real · 593 AI-generated

0.93
F1, AI-generated video
balanced group: 477 real · 455 AI

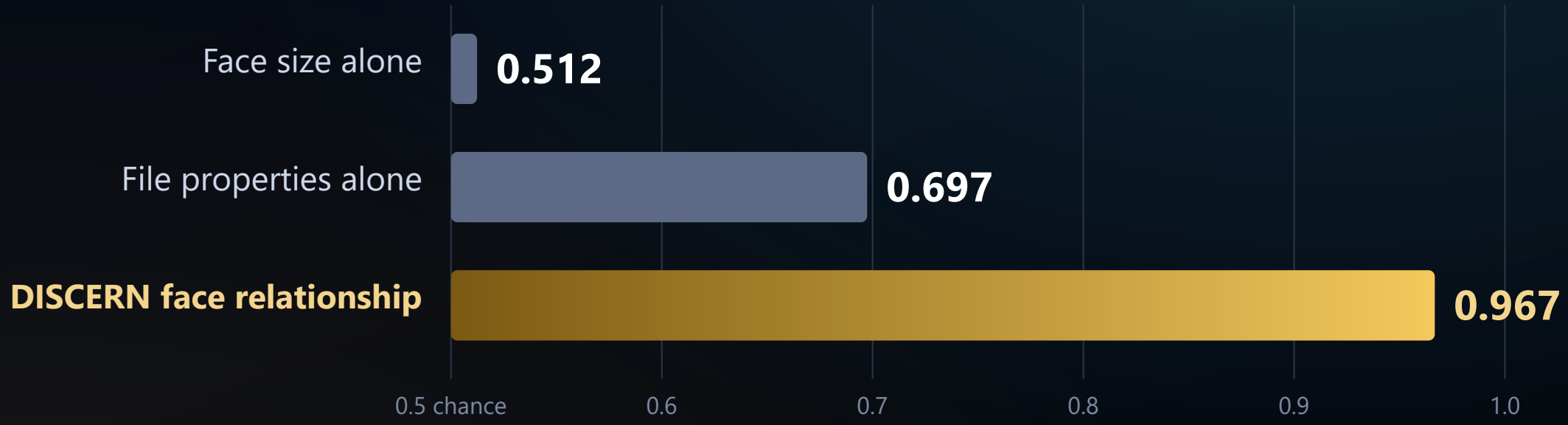
0.93
F1, real video
same group

11
video generators
including Sora, Kling, Luma, Runway Gen-3

AIGVDBench (CVPR 2026): 4,031 real videos and 2,000 from each of 11 generators, 26,031 in all. Scored: every clip in which one face is held for four seconds. Real and AI clips are matched in frame rate, length, resolution and encoding. Gradient boosting, five-fold cross-validation grouped by source video. AUC 95% interval 0.967–0.981; balanced-group AUC 0.973.

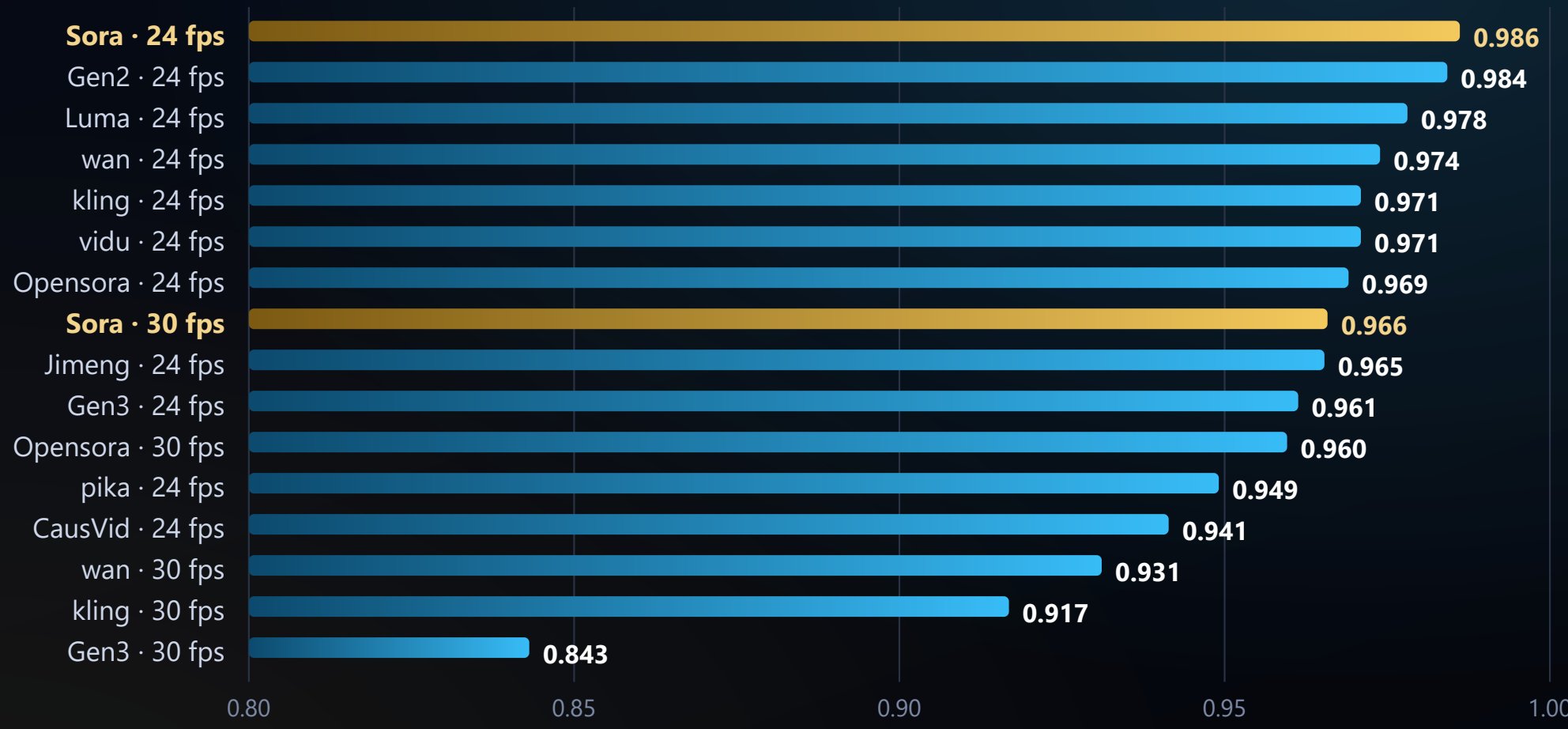
VIDEO · NOT THE ENCODING. NOT THE FACE SIZE.

Same encoding, same face size: the living-face signal stays at 0.967.



24 fps group, every clip re-encoded identically (96 frames, 512×512, 1 Mbit/s) and limited to faces of matching size (106–230 px): 322 real, 247 AI-generated. Gradient boosting, five-fold cross-validation grouped by source video. Face relationship AUC 95% interval 0.952–0.979. A follow-up analysis on the same data.

Each generator held out of training in turn. Sora: 0.986.



AUC on the held-out generator; the model never saw a single clip from it. Mean across all 16 held-out tests 0.954; 24 fps group mean 0.968.

VIDEO · FACE SWAPS

One model for both kinds of forgery: generated video and face swaps.

0.795

AUC, each face swap against its own unswapped original

0.972

AUC on generated video, same model

43

swap–original pairs
5 source faces · DeepFaceLab



Face swaps built in-house with DeepFaceLab from AIGVDBench footage. Each swap and its original share the same frames, cut and encoded identically (96 frames, 512×512, 1 Mbit/s); they differ only in the face. One gradient-boosting model trained on AIGVDBench and the swaps together, five-fold cross-validation with every clip from the same source video in the same fold. Swap AUC 95% interval 0.740–0.853.

Four signals a generator leaves behind, each one measured, none of them learned.

0.87

The face edge trembles

On generated faces, the boundary of the face jitters against its own interior. A living face moves as one piece.

0.80

Skin that never changes

Generated skin texture stays glued in place frame after frame. Living skin texture shifts with every movement.

0.69

Too clean to be real

Real video carries fine temporal grain from light and sensor. Generated video is smoother than the world.

0.66

Light that does not travel

On a real face, highlights slide across the skin as the head turns. On a generated face they sit still.

Number shown: each signal on its own, untrained, in the 24 fps group (AUC, distance from chance). Every signal points the same way in both frame-rate groups.

VOICE · CLONED AND SYNTHETIC SPEECH

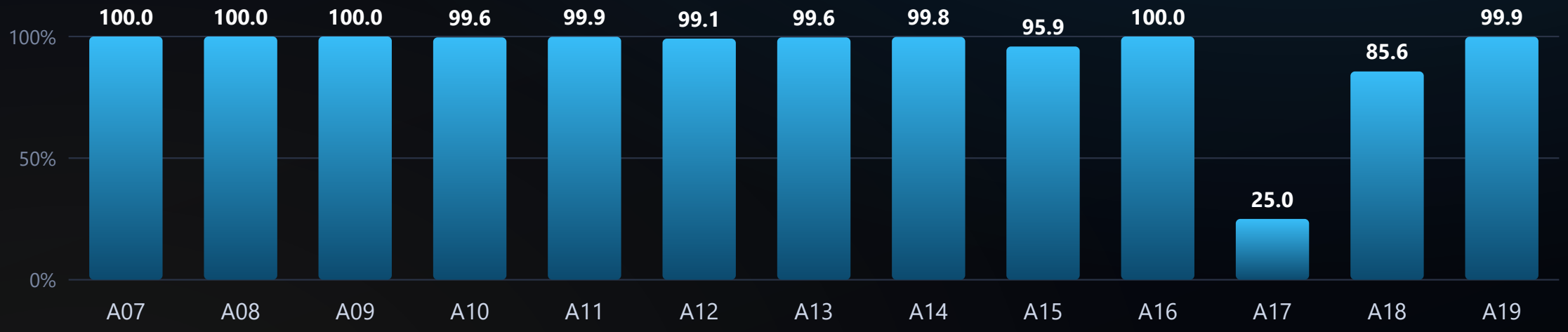
The full ASVspoof 2019 evaluation: 71,237 utterances, none left out.

0.9747
ROC-AUC

7.36%
Equal error rate

0.9576
F1

11 / 13
attack systems caught at 95.9% or higher



Catch rate per spoofing system, A07–A19, at the equal-error threshold. Every spoofed utterance in the evaluation partition is scored; bonafide speech cleared: 6,814 / 7,355.

AT A GLANCE

Two channels, both kinds of video forgery. Every number from a complete run.

CHANNEL	BENCHMARK	RESULT	SCOPE
Video	AIGVDBench, CVPR 2026 · CC BY 4.0	**AUC 0.974**	1,312 real and 593 AI clips from 11 generators · F1 0.93 AI / 0.93 real
Video, unseen generator	Same, each generator held out	**Sora 0.986**	Mean 0.954 across 16 held-out tests
Video, face swaps	In-house DeepFaceLab swaps, AIGVDBench footage	**AUC 0.795**	43 swaps, each against its own original · one model with generated video (0.972)
Voice	ASVspoof 2019 LA · ODC-BY	**AUC 0.9747**	All 71,237 evaluation utterances · EER 7.36% · F1 0.9576

AIGVDBench: L. Ma et al., "Your One-Stop Solution for AI-Generated Video Detection," CVPR 2026, CC BY 4.0.
ASVspoof 2019: M. Todisco, X. Wang, V. Vestman, M. Sahidullah, H. Delgado, A. Nautsch, J. Yamagishi, N. Evans, T. Kinnunen, K. A. Lee, "ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection," Proc. Interspeech 2019, ODC-BY.

Paul Samuel Guarino · 41.176hz@gmail.com