moveris™

WHITE PAPER

Protecting Online Meetings from Deepfakes, Synthetic Identities, and Impersonation Fraud

Why Human Physiology is the Next Line of Defense

> On a Friday afternoon in early 2024, a finance employee at the engineering firm Arup joined a video call with several colleagues, including the company's Chief Financial Officer. Everyone on the call looked and sounded right. Over the course of the meeting, the employee was instructed to move money for a confidential transaction, and did, wiring $25 million across fifteen transactions to five Hong Kong bank accounts. Every person on the call except the victim was a deepfake.[1]

The Arup case is the clearest and most well-known illustration to date of a structural problem. The video call, long treated as one of the more trustworthy channels in virtual communication precisely because it lets people see and hear each other, is now an attack surface. Generative AI has made real-time audio and video impersonation cheap, fast, and convincing enough to defeat both human judgement and most of the automated defenses built to catch it.

This paper surveys the solutions that have emerged to address this problem, including cryptographic identity credentials, continuous voice and video analysis, forensic artifact detection, and challenge-response liveness checks. We then argue that each of these types of solutions share a common vulnerability in that they evaluate appearance, voice, or a one-time credential, all of which generative models are explicitly optimized to reproduce. We then describe Moveris' alternative approach, which verifies a participant is a living human by measuring involuntary physiological processes, such as blood flow and facial micro-expression, and checking whether they cohere with each other in real time. That coherence is a byproduct of biology, not appearance, and is what current generative systems are not built to simulate.

The Threat: Video Calls Have Become an Attack Surface

Video conferencing assumes that seeing and hearing a colleague in real time is sufficient proof of who they are. Three trends have quietly invalidated that assumption over the past three years.

First, generative AI has made live impersonation cheap and fast. For example, voice cloning models can now produce a convincing synthetic voice from as little as four seconds of reference audio. Real-time face-swap and avatar-rendering tools, originally built for entertainment and virtual production, run live on consumer-grade hardware with latency low enough to sustain a natural back-and-forth conversation.

Second, the losses are already material and growing quickly.[2] Fraud attempts leveraging deepfake content have climbed more than 2,000% since 2025. AI-enabled fraud losses are projected to reach approximately $40 billion by 2027, up from $12.3 billion in 2023. Synthetic identity fraud alone costs U.S. financial institutions an estimated $30-35 billion annually, with lenders carrying over $3.3 billion in exposure to synthetic identities in new accounts. And 64% of fraud

professionals now name deepfakes as a top fraud concern, while 62% of banks identify digital onboarding as their point of highest synthetic-identity exposure.

Third, deepfake detection is losing ground to generation. Independent evaluation of deepfake voice detectors has found accuracy as low as 38.2% against current-generation cloning tools.[3] This is not a temporary gap that better training will close; it is the structural signature of an adversarial arms race in which every improvement to a detector is, by design, a training signal for the next generation of generator.

Why Meetings Are a Uniquely Vulnerable Channel

It has long been thought that a live video call could not be compromised, and organizations have built verification habits accordingly ("let's hop on a call to confirm"). The Arup incident shows that way of thinking breaking down in real time, in front of multiple participants, none of whom questioned what they were seeing.

Meetings are also harder to protect than static content. A photo or a recorded video can be authenticated after the fact, but a live meeting has no “after the fact.” Any verification has to happen continuously, in real time, without breaking the flow of a natural conversation, and without asking participants to interrupt the meeting to complete a challenge. This is where most existing solutions run into trouble.

The Current Solution Landscape

Meaningful engineering effort has gone into defending video calls over the past two years. Broadly, the field has converged on three categories of defense:

1. **Cryptographic proof-of-personhood**

   A user enrolls once at an in-person iris-scanning device and receives a portable credential tied to a unique human rather than to a name or a face template. When joining a meeting, the app checks a live selfie and the current video frame against the original enrollment and issues a verification badge to other participants, without sharing biometric data.

2. **Continuous multimodal identity verification**

   The system passively analyzes a participant's voice and video throughout a meeting for the signatures of cloning, replay, or face-swapping, and correlates that with contextual risk signals such as VPN use or a mismatched claimed location.

3. **Real-time forensic artifact detection**

   Bots join a call and run multimodal neural networks against the raw video and audio in real time, scoring each participant's likelihood of being synthetic and alerting a host or security team when the score crosses a threshold. Another related solution projects a randomized sequence of colors onto the screen and verifies that the reflected light appears correctly on the user's face.

The Fundamental Limitation: Appearance Can Be Rendered

The issue with each of these solutions is that appearance-based and voice-based detection is a problem played against an adversary whose entire job is to defeat that problem. Every time a detector improves, it does so by learning the visual or acoustic signature of the generators it has already seen. And every improvement to detection is, in turn, a specification for the next generator to route around.

> *"Every improvement to detection is, in turn, a specification for the next generator to route around."*

The Moveris Approach: Physiology-Based Human Verification

Moveris starts from a different premise. Instead of asking whether a face or voice looks convincing, our technology asks whether the underlying biological processes that keep every living person's body running are actually present, and whether they behave the way real physiology should behave in real time and in response to what is actually happening on the call. To do this, Moveris measures several physiological channels simultaneously using nothing more than the existing camera feed already present in a video call, and then analyzes how the signals synchronize with each other in the correct proportions, in real time, and in response to a conversation whose content the attacker does not fully control. Thus, multimodal physiological coherence, not any single signal, is the defense.

Why This Is a Different Category of Defense, Not a Better Version of the Same One

It's worth noting what changes with this approach. Physiological coherence checking evaluates real-time behavior of independent biological subsystems that exist to keep a person alive, not to make them look convincing on camera, something the attacker does not control. This reframes the underlying arms race. Instead of a contest over visual or acoustic realism, the bar becomes simulating human biology itself, correctly, across multiple systems, synchronized to a live and unscripted conversation. That is a more difficult and durable bar to clear, and it is one that does not reset every time a new generative model ships. MoverisMeeting is an API built to run passively alongside an existing video call rather than in place of it. It processes the video streams already flowing through the meeting, extracts the physiological signals, and returns a human likelihood score that a host, a security team, or an automated policy engine can act on without interrupting the conversation.

Conclusion

The Arup incident was not a failure of vigilance. Generative AI has simply made looking and listening no longer sufficient, and the industry's response so far has largely been to build better versions of looking and listening. But a more lasting answer is to stop asking whether a call participant looks or sounds human, and start asking whether their body is behaving like one.

[1] World Economic Forum, “Cybercrime: Lessons learned from a $25m deepfake attack.” https://www.weforum.org/stories/2025/02/deepfake-ai-cybercrime-arup/

[2] BIIA.com, “Synthetic Identity Fraud Statistics 2026: Hard Numbers, Big Threats.” https://www.biia.com/synthetic-identity-fraud-statistics-2026-hard-numbers-big-threats/

[3] Ibid.