Protect your organization from deepfake scams

Cisco offers security that ensures authentic collaboration for our customers

A growing number of businesses are concerned that a deepfake scammer might infiltrate their organization, for example, by imitating an executive and convincing someone in their finance department to make a large transaction. Fortunately, as a Cisco customer, you can protect yourself and your organization against this kind of scenario. Read on for a summary of how Cisco provides protection against deepfake-enabled scams and how you can stay safe.

Key takeaways:

- The best option for Cisco customers is to ensure all Webex meeting participants have authenticated identities and all meetings have proper access controls.
- To remove the risk from deepfakes entirely, organizations will ultimately need to update their processes so that they are not susceptible to social engineering.
- Webex provides several tools that customers and Webex meeting participants can use to make sure they're safe from deepfake accounts, and more are on the way.

How is your organization at risk?

Deepfakes are a dark byproduct of generative AI, one of the largest technology transformations the world faces today. The generative AI revolution has yielded algorithms that can reproduce a person's likeness and voice with very high fidelity. Certain companies offer voice cloning, while others offer product lines dedicated to video manipulation.

These algorithms have several beneficial uses (for example, creating audiobooks from text or improving one's appearance on a video call), but they can also be used to create a person's voice or image without their consent. These algorithms are also getting better and better over time.

Many organizations are already frequently targeted by scams where bad actors impersonate vendors or other employees, but deepfake technology makes these attacks much more powerful: An employee might flag as suspicious an email from their CEO instructing them to make a payment, but would they say "No" to a CEO directly on a video call?

Existing methods of protection

There are three basic approaches to countering deepfake scams:

Deepfake detection: Algorithms estimate how likely a voice or video sample is to be a deepfake. If these algorithms were integrated into collaboration products, they could provide a user with a warning that the person they are talking to might not be authentic.

Authentication and Access Control: There are several mature technologies for verifying the identities of the people in a meeting and making sure that only the right people are allowed into a meeting. These technologies make it impossible to perform a deepfake scam without breaking into a user's account — a much higher bar than just running a deepfake algorithm.

Process adaptation: Deepfake scams exploit personal interactions within business processes. For example, an employee might pay an invoice based on the verbal approval of a manager. If these processes can incorporate technical controls (e.g., requiring the manager to log in to a system to approve a payment), then they will be immune from deepfake attacks.

Benefits and limitations of existing methods

The three approaches described above complement one another. They represent three trade-offs between ease of deployment and the strength of the protection against deepfakes.

Approach	What a scammer must do to successfully attack	What a user must do to successfully defend	
No Mitigation	Make a deepfake of someone the victim trusts	Recognize that the voice / image is not authentic	
Deepfake Detection	Make a deepfake that bypasses the detection algorithm **or** Convince victim to ignore detection indicator	Recognize detection indicator **and** Insist on enforcement	
Identity	Break into the deepfaked user's account	Recognize and enforce authentication indicator **or** Enable access controls	
Process Adaptation	Break into the deepfaked user's account	(Nothing)	

Figure 1: Approaches to mitigating the risk of deepfake scams

Deepfake Detection

One advantage of deepfake detection is that it can be unilaterally included in a product without customers or users having to take any action to set it up. However, it's difficult for vendors to ensure that these algorithms hold up under adversarial conditions, and even the best algorithms have a degree of inherent uncertainty.

Even when equipped with warnings from a deepfake detection system, users still have to notice and respond to the warnings — and do so in situations when they are already under stress.

Imagine a scenario where an employee is speaking to a deepfake scammer's imitation of their CEO: The employee receives a warning of a potential deepfake, so the deepfake CEO says, "You can ignore that warning. Something about my camera here always sets that off." At this point, the safety of the organization depends on that employee's willingness to contradict someone they believe to be the CEO.

Authentication and Access Control

Authentication is a mature technology that provides high assurance: To spoof an authenticated identity, a scammer needs to break into the victim's account or convince the victim to trust a fake identity — a much higher bar than just making a deepfake.
Let's go back to our employee talking to the deepfake CEO. If the employee can verify that the person they're talking to is a legitimate, authenticated user in a specific organization, that's a clear signal that the person they're talking to is unlikely to be a deepfake.

Of course, the employee still must be paying attention to the authentication indicators in the first place, and they must be able to recognize when the person they're talking to is not from the organization they expect. Again, social engineering presents a risk: Imagine that the deepfake CEO has an identity from company-elt.com instead of company.com and says, "Oh, that's a domain that we in the Executive Leadership Team use."

Authentication technologies work well with other Webex security technologies, such as end-to-end encryption and access controls. With end-to-end identity technologies, users receive direct verification of a remote party's identity, so that not even Cisco can replace the user with a deepfake.

Access control technologies automate the checks Alice was doing above, so that the risk of social engineering goes away. If Alice had actually invited the CEO to a meeting, using the CEO's real email address, and configured the meeting so that only invited participants could join, then the deepfake user from a different domain wouldn't have been able to join the meeting at all.

Maintaining a robust authentication system entails some work for an organization, but this is work many organizations already do to support features like Single sign-on. Webex and similar apps already integrate well with customers' authentication systems. Cisco security products like Duo and Identity Intelligence can provide extra layers of security.

Process Adaptation

This approach is highly specific to an organization's processes and culture. Structuring business processes so they rely on technical controls, rather than humans recognizing one another, provides the best possible level of security against deepfake scams, because it removes the underlying vulnerability that deepfake scams exploit: humans' ability to be fooled. However, implementing the needed technical controls requires examining each of an organization's processes individually and often requires upgrades to the information systems that facilitate these processes.

The adaptations needed here are similar to what companies have done in the past to avoid scams in other media. In the days before computers, an accounts payable department wouldn't make a payment just because someone claiming to be a vendor called them or based on the verbal approval of an executive. They would insist on a paper invoice on the vendor's letterhead, with physical signatures from the relevant approvers, something much harder to forge. For similar reasons, employees are commonly trained to be suspicious of requests they receive over email to stay safe from email-based scams.

What we need now is the electronic equivalents of those hard-to-forge documents. These will take different forms for different organizations, but they are typically represented by the actions of authenticated users in the electronic systems that manage business processes, e.g., finance and accounting systems. If a CFO wants a payment made, they should have to log in to the finance system and approve it, not just call someone who could mistake them for a deepfake.

What you can do to stay safe

Given the trade-offs of the tools mentioned, there are several steps you can take today to help keep your organization safe from deepfake scams:

- Protect your meetings.
    - The best situation is when everyone in a Webex meeting is logged in. In this situation, a deepfake scammer can't join the meeting unless they hack a legitimate user's account — a much higher bar than just making a deepfake. You can require sign-in for your organization's Webex meetings in Control Hub.
    - Follow our best practices for secure meetings to configure appropriate security controls for your organization's Webex meetings.
    - Watch out for "unverified" participants and call-in users.

    - Lock your meetings to make sure that only authorized participants can join, and use the information Webex provides when deciding who to let in.
    - Use tools like space meetings to make it easy for the right people to join while still locking out the bad guys.
- Protect your logins.
    - Use single sign-on (SSO) instead of application-specific logins.
    - Use phishing-proof, passwordless authentication.
    - Use Identity Intelligence to detect misuse of users' accounts.
- Help your employees become fraud-proof.
    - Make users aware of the risks from deepfakes.
    - Call back: If someone calls you and asks you to do something strange or suspicious, the simplest defense is to call them back. This implicitly authenticates their identity, since only devices where the user is logged in will ring.
    - Emphasize the importance of following business processes and empower employees to push back when superiors seem to be asking for suspicious requests.