Why Your Family and Coworkers Need a Safe Word in the Age of AI

Scammers always rely on deception to trick people out of their money, but the sudden wide availability of artificial intelligence (AI) has increased the convincingness of their schemes.

AI can be used to clone voices and even create deepfake videos, so we all need to be vigilant about any incoming communication, from emails and texts to calls and FaceTimes. These tools allow criminals to replicate the voices of family members, coworkers, and even government officials with eerie accuracy. All that is required is a short audio or video clip.

However, by creating a safe word as a preventative measure, you can stop AI-empowered scams in their tracks.

How audio cloning scams and deepfakes work

A common AI scam starts with a distressed call or voice message from a loved one – or, at least, someone speaking in your loved one's voice. Video calls with a "deepfake" can also happen, meaning the scammers use AI to replicate both video and audio of one of your contacts. The voice on the other end might say they've been in an accident, are in legal trouble, or need urgent financial help. Because the voice sounds exactly like your child, spouse, or boss, your instinct is to react immediately and not think before you take action. In just a few moments, you've fallen victim to the scam by taking actions like wiring money or sharing sensitive data.

How scammers create deepfakes

With just a short audio sample—taken from social media, voicemail recordings, or even a quick phone call—criminals can generate fake messages that sound exactly like someone you trust. The same works for video deepfakes. There might be a significant amount of video footage of yourself posted online. Today's AI systems only need around 30 seconds of audio and video data to create convincing clones.

Establish a safe word to keep your family safe

A safe word is a pre-agreed code word or phrase only you and your trusted group know. If you ever receive an urgent call or message, asking for the safe word is a quick way to verify the person's identity.

Who should have a safe word?

Safe words aren't just for families, they can be useful in many settings. Here are some groups that should consider setting up a safe word system:

- **Families:** Parents and children can use safe words to verify identity in emergency situations or potential kidnapping scams (commonly called grandparent scams).
- **Coworkers and teams:** Businesses can establish safe words to confirm financial transactions, sensitive information requests, or urgent company-wide messages.
- **Close friends:** Friends can also have a verification system in place, especially if one is traveling or in a vulnerable situation.
- **Elderly adults and their caregivers:** Scammers often target older adults, so having a safe word between an elderly person and their trusted family or caregivers is critical.
- **Online communities and groups:** Any organization that relies on virtual communication can benefit from a safe word to prevent impersonation, especially if the group has access to sensitive personal data.

How to set up a safe word system to defend against AI

1. **Make it unique.** Pick a word, phrase, or question that wouldn't be easy for an outsider to guess. Avoid common passwords, birthdays, or pet names. Inside jokes are ok if they are indeed impossible for an outside party to understand.
2. **Keep it private.** Share your safe word in person or via a secure, encrypted channel. Never post it online or discuss it in public settings. You can store a safe word in your encrypted password manager vault. If anyone posts a hint for the safe word online, even if it wasn't intentional, change the safe word.
3. **Use multiple safe words, but make them each special.** While families should have a safe word, workplaces can benefit, too. Coworkers and teams should establish verification codes for sensitive or financial requests. Don't use the same safe word among multiple groups, however.
4. **Practice using the safe word.** Like any security measure, a safe word only works if people remember to use it. Occasionally test your safe word system to make sure everyone involved recalls the procedure. Testing it doesn't have to be scary, it could be a regular game.

Extra verification steps

While a safe word is an excellent first line of defense, here are additional ways to verify someone's identity:

- **Call back using a trusted number:** Hang up and call the person back using a known number stored in your contacts. Hackers can spoof phone numbers so you should call back from the phone number you have stored, even if the person on the other end begs you to stay on the line.
- **Video verification:** Ask to switch to a video call if possible. While video deepfakes can happen, it is less likely that a scammer has both a video deepfake and an audio clone created. Even still, ask for your safe word.
- **Ask personal questions:** Have a secondary verification method, like a question only the real person would know the answer to. This can help you verify the person beyond a single safe word or phrase.

Don't panic and Stay Safe Online

Deepfake scams are sophisticated, but you can keep your head calm and take precautions like safe words. A safe word is a simple, powerful, and very human tool anyone can use to verify identity and prevent scams. Whether with your family, friends, or coworkers, take the time today to establish a safe word because it's our best defense against AI-powered scams.

The information provided in the MS-ISAC Monthly Cybersecurity Tips Newsletter is intended to increase the security awareness of an organization's end users and to help them behave in a more secure manner within their work environment. While some of the tips may relate to maintaining a home computer, the increased awareness is intended to help improve the organization's overall cyber security posture. This is especially critical if employees access their work network from their home computer. Organizations have permission and are encouraged to brand and redistribute this newsletter in whole for educational, non-commercial purposes.

Disclaimer: These links are provided because they have information that may be useful. The Center for Internet Security (CIS) does not warrant the accuracy of any information contained in the links and neither endorses nor intends to promote the advertising of the resources listed herein. The opinions and statements contained in such resources are those of the author(s) and do not necessarily represent the opinions of CIS.