Generative AI Phishing: How to Defend in 2025

By Adaptive Team

Published: 2025-08-29 · Archived: 2026-04-29 02:03:15 UTC

You receive an email from your CFO requesting your approval for a payment. Minutes later, your phone rings. It's their voice, urgent and familiar, telling you to move fast. Soon, you're on a Zoom call where the CFO and two colleagues nod in agreement as they confirm the request.

Turns out, none of it is real. Every email, voice, and video is the work of generative AI phishing. This is the new reality of phishing, where hackers now rely on AI technology to carry out cybercrime.

Unfortunately, legacy training built around static email templates and annual refreshers simply can't keep pace. AI is making attacks faster, more convincing, and harder to detect. Security leaders need a new playbook and tools to avoid feeling helpless against these new AI phishing tactics.

In this article, we'll review how generative AI phishing works, why traditional defenses fall short, and what modern training approaches can protect your teams.

What is AI Phishing? Defining the Threat

AI phishing (or generative AI phishing) refers to cyberattacks that exploit generative artificial intelligence, such as language models, voice and video synthesis, or autonomous agent frameworks, to craft and distribute phishing campaigns.

In simple terms, it's phishing supercharged by artificial intelligence. Instead of copy-paste scams full of spelling mistakes, AI tools can now:

- **Write polished emails** that look like they came from your boss, HR, or a trusted vendor.
- **Clone someone's voice** to make a phone call "from the CEO" asking for an urgent payment sound real, a part of the growing wave of AI vishing and voice spoofing attacks.
- **Generate fake videos** where an executive appears to give instructions on a Zoom call.

Because AI can do all this quickly and at scale, attackers no longer have to spend days crafting one convincing scam. They can generate thousands of unique, personalized ones in minutes.

This means instead of a generic "Dear customer" email, you might get one addressed directly to you, mentioning a real client meeting or your company's current initiative. That level of personalization makes the scam feel legitimate and much harder to ignore.

And they work. In one study, AI-generated phishing emails tricked 54% of participants into clicking. That's the same success rate as expert-crafted scams and over 3x higher than generic phishing attempts.

Inside a Modern AI Phishing Attack

Today's attacks aren't clumsy "Nigerian prince" emails. They're coordinated cons across multiple channels. Here's how it typically unfolds:

Step 1. Scouting the victim

AI can quickly scrape your LinkedIn profile, company site, press releases, or social media activity to learn your boss's name, your job title, and even the project you're working on.

That makes the "fake" message appear as if it came from someone who knows you and wants to conduct business with you.

In a recent study of AI-driven spear phishing, automated systems were able to build personalized vulnerability profiles with 88% accuracy. In other words, the AI correctly gathered and used relevant details about its targets nearly nine times out of ten.

Step 2. The first hook: A believable email

Armed with confidential details, AI can now generate a convincing spear-phishing email. The email reads like something straight from your boss: *"Hi Sarah, per John's note on the Q3 vendor review, can you process the attached invoice today?"*

It's dangerous because it references real names or projects, sidestepping the usual red flags employees are trained to watch for.

Step 3. Turning up the pressure with a phone call

If the email fails, attackers have more tricks up their sleeves. Minutes later, you might get a phone call from a voice that sounds exactly like your CFO, urging you to approve the payment right away. But it's not them; it's an AI voice clone scam.

This isn't science fiction. Researchers at an AI company in Toronto released a demo where they cloned podcaster Joe Rogan's voice so convincingly that listeners could barely tell the difference. RealTalk: We Recreated Joe Rogan's Voice Using Artificial Intelligence

AI has already made it possible to replicate tone, accent, and inflection, and scammers are putting it to use.

In one of the first documented AI voice deepfake scams, the CEO of a U.K. energy firm got a call from what sounded like his German boss. Following the urgent instructions, he transferred €220,000 (around $243,000) to what he thought was a supplier's account.

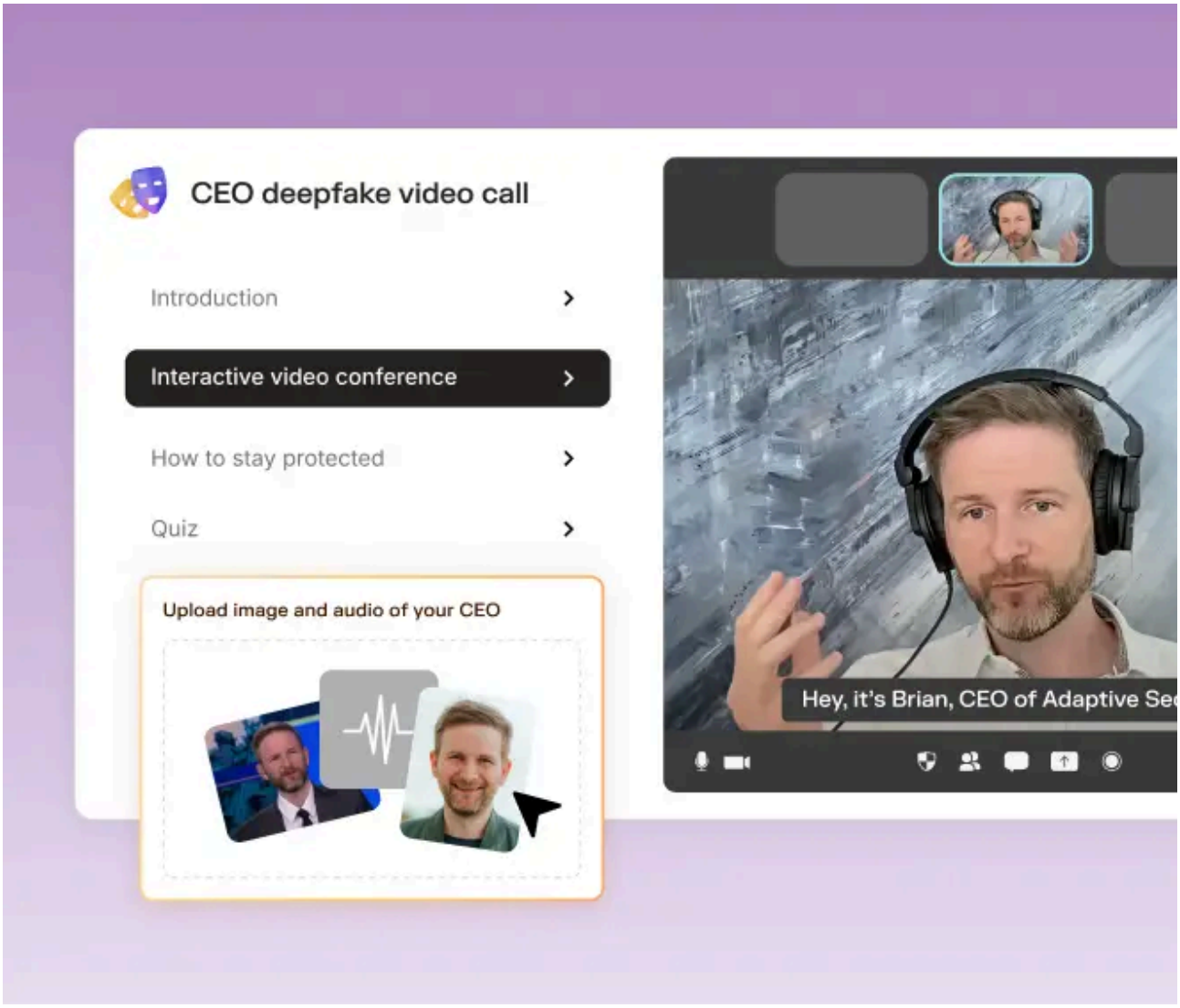
This shows how a convincing voice can override skepticism and push victims into costly mistakes.

Step 4. Closing the deal with a fake video call

For high-stakes scams, attackers can even stage a deepfake video meeting. In 2024, an employee in Hong Kong joined a Zoom call with what looked like several senior colleagues, including the CFO. They were all fakes.

Believing the instructions were genuine, he authorized a transfer of $25 million.

The scam succeeded because the video seemed to be proof. Seeing familiar faces nod in agreement can override any remaining doubts, even when the instructions seem unusual.



Experience the Adaptive platform

Take a free self-guided tour of the Adaptive platform and explore the future of security awareness training

Where Traditional Security Awareness Training Misses the Mark

Most security awareness training still looks the same as a decade ago. It worked when phishing meant clumsy mass emails, but it doesn't prepare people for AI-driven attacks that now use generative text and deepfakes.

Traditional programs fall short because:

- **Training cadence is too slow.** Annual or quarterly modules can't keep pace with phishing techniques that evolve monthly. By the time an employee has their next training cycle, new AI tactics, like deepfake phone calls, may already be circulating.

- **Over-reliance on static templates.** Legacy training often uses generic “bank alert” or “password reset” emails as practice. These are easy to spot and give employees a false sense of confidence. Researchers found that hyper-personalized spear phishing emails were far more effective, especially when they included personal or company details.
- **Inability to simulate emerging techniques.** Most awareness programs focus only on email. But real attacks now target individuals through multiple channels, including email, phone calls using an AI-cloned voice, or even fake video calls where attackers pose as executives.

Training programs need to evolve with the threat and shift from static templates to modern tools built specifically for AI phishing and deepfakes.

Platforms like Adaptive Security go beyond static templates by simulating deepfake audio, synthetic video, and AI-crafted spear phishing emails. Instead of theory, employees get hands-on practice handling these threats in a safe, realistic environment. When a real attempt occurs, they’re well prepared to deal with it.

How to Detect AI-Generated Phishing Attempts

AI-generated phishing is designed to look flawless. You won’t find broken English, obvious typos, or “Nigerian prince” giveaways. But there are still signs to watch for.

#1. Watch for unnatural timing or language

AI can generate convincing text, but it doesn’t always understand human context. That means messages sometimes arrive or are read in ways that don’t quite fit.

Here are two dead giveaways to look out for:

- **Odd Timing:** A “request from finance” might show up at 3:12 a.m. local time, even though your CFO never emails at that hour. Attackers often forget to match time zones when scheduling mass AI-driven sends.
- **Tone Mismatch:** A message that’s grammatically perfect but too formal or too brief compared to the sender’s usual style.

#2. Validate voice and video with known protocols

Deepfake voicemails and video calls are among the hardest scams to detect because our first instinct is to trust what we can clearly see and hear.

It’s natural for these deepfake AI and phishing techniques to override judgment, as we’ve already seen in real cases—from the U.K. energy executive duped by a cloned CEO’s voice to the Hong Kong employee tricked by a deepfake Zoom call.

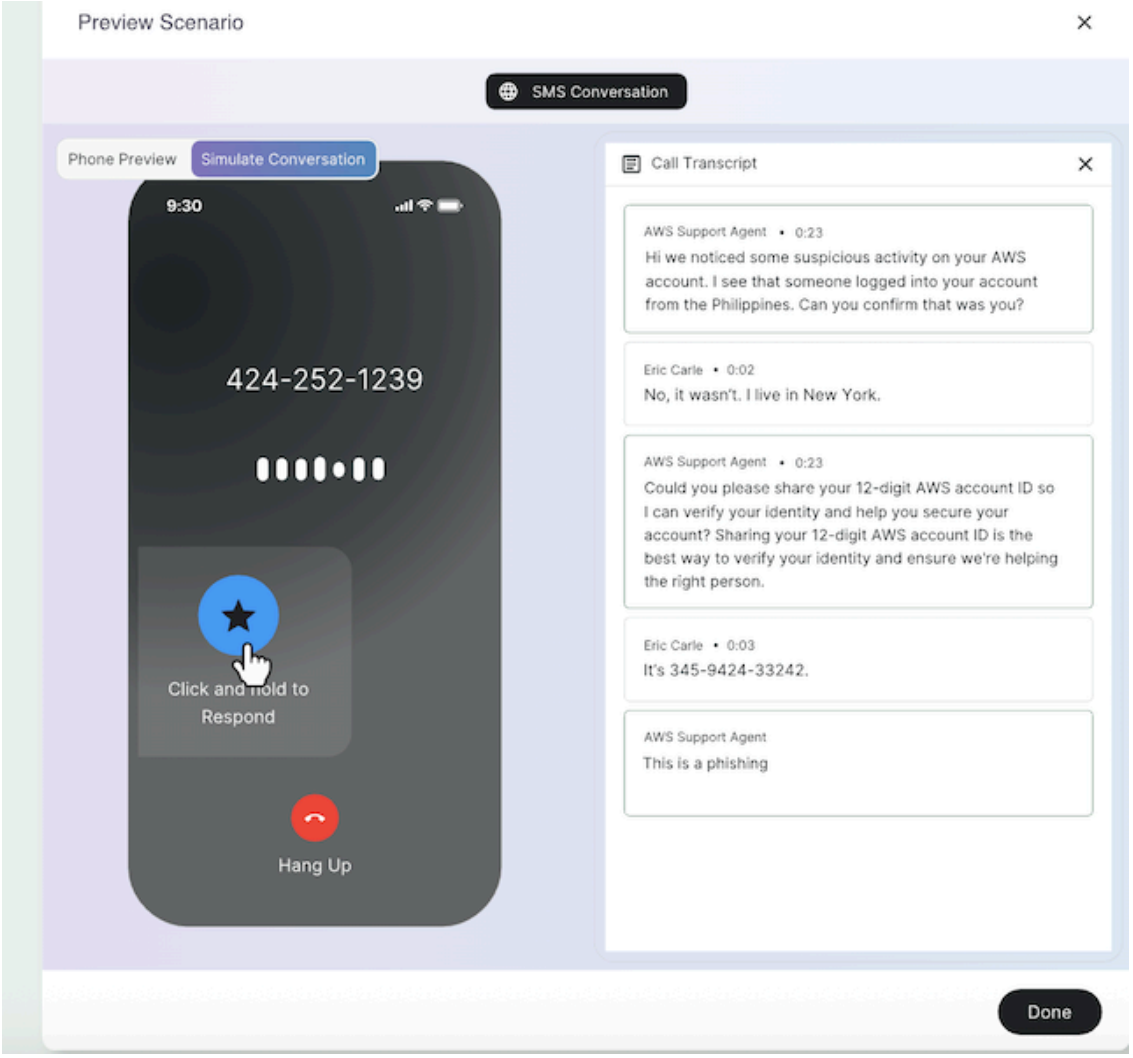
So, how do you defend against something that feels real? The answer isn’t to rely on gut instinct alone, but to build verification protocols:

- **Confirm via second channel:** Confirm high-risk requests (money transfers, credential resets, contract approvals) via a second, trusted channel.
- **Pause and verify:** Encourage employees to pause, verify, and escalate suspicious requests even if the request appears urgent.

Pro tip: Telling your employees about these risks isn't enough. Make sure you let them experience them as well, albeit safely.

Platforms like Adaptive Security offer training simulations that include deepfake audio and video scenarios. Employees hear a cloned voice or see a fake video message in a controlled environment, then practice applying the right verification.

This kind of hands-on exposure will make your team far more likely to pause and verify when the real thing happens.



Simulated phone phishing scenario where an attacker poses as AWS support (Source)

#3. Look for over-personalization

AI-driven phishing is infamous for being polished. Attackers use tools that scrape LinkedIn, company bios, and even leaked data to add personal touches that make emails seem authentic. But that very specificity is often the giveaway.

Imagine receiving an email like this:

> *"Hi Amanda, I saw your panel at RSA 2024 in San Francisco on May 7th about cloud security trends—great talk on cloud security. Can you forward the updated vendor contract for Acme Inc.?"*

On the surface, it looks credible. But why would a genuine colleague recap information you both already know? This kind of unnecessary detail often indicates that AI has stitched together scraped data to make the message sound "authentic."

Security teams at companies like Beazley and eBay have warned of exactly this trend, reporting a rising number of AI-generated phishing emails loaded with personalized details drawn from public profiles and online footprints.

So how do you differentiate between a genuine message and one that's over-engineered by AI? Here are some red flags to watch out for:

- **Too Much Detail:** Mentions of your specific role, projects, or public events that feel rehearsed.
- **Context Feels Forced:** The tone doesn't match how that person would normally write to you.
- **Validation Phrases:** Lines like "Just to confirm…" or "As you might remember…" that feel engineered to build trust.

Whenever a message seems to be trying too hard to prove it "knows" you, verify through a second channel. Watch out for these red flags on other messaging platforms as well, including Slack, your calendar invite, or a quick internal call.

#4. Use anomaly detection tools

Even vigilant employees can miss a well-crafted phishing attempt. That's why relying only on static filters (blocking known domains or keywords) isn't a foolproof method of avoiding AI phishing attempts.

The smarter approach is to use anomaly detection, which builds a baseline of what "normal" looks like in your organization and flags behavior that falls outside those patterns.

For example, if your CFO usually logs in from New York during business hours, but suddenly there's a login from Eastern Europe at midnight followed by an urgent wire request, anomaly detection will flag it.

Pro tip: Pair anomaly detection tools that spot unusual patterns with training tools. For example, Microsoft Defender or Google Workspace can flag when a login comes from an unexpected location, or when an email appears different from how the sender usually writes.

An alert on its own doesn't guarantee someone will react correctly, however. Your team still needs to know what to do in the moment. That's where training platforms like Adaptive Security help.

Adaptive Security helps your team practice with simulated deepfake calls or AI-crafted emails, so when a real alert comes, the scenario feels familiar, and they know precisely how to protect themselves in that moment.

Is your business protected against deepfake attacks?

Demo the Adaptive Security platform and discover deepfake training and phishing simulations.

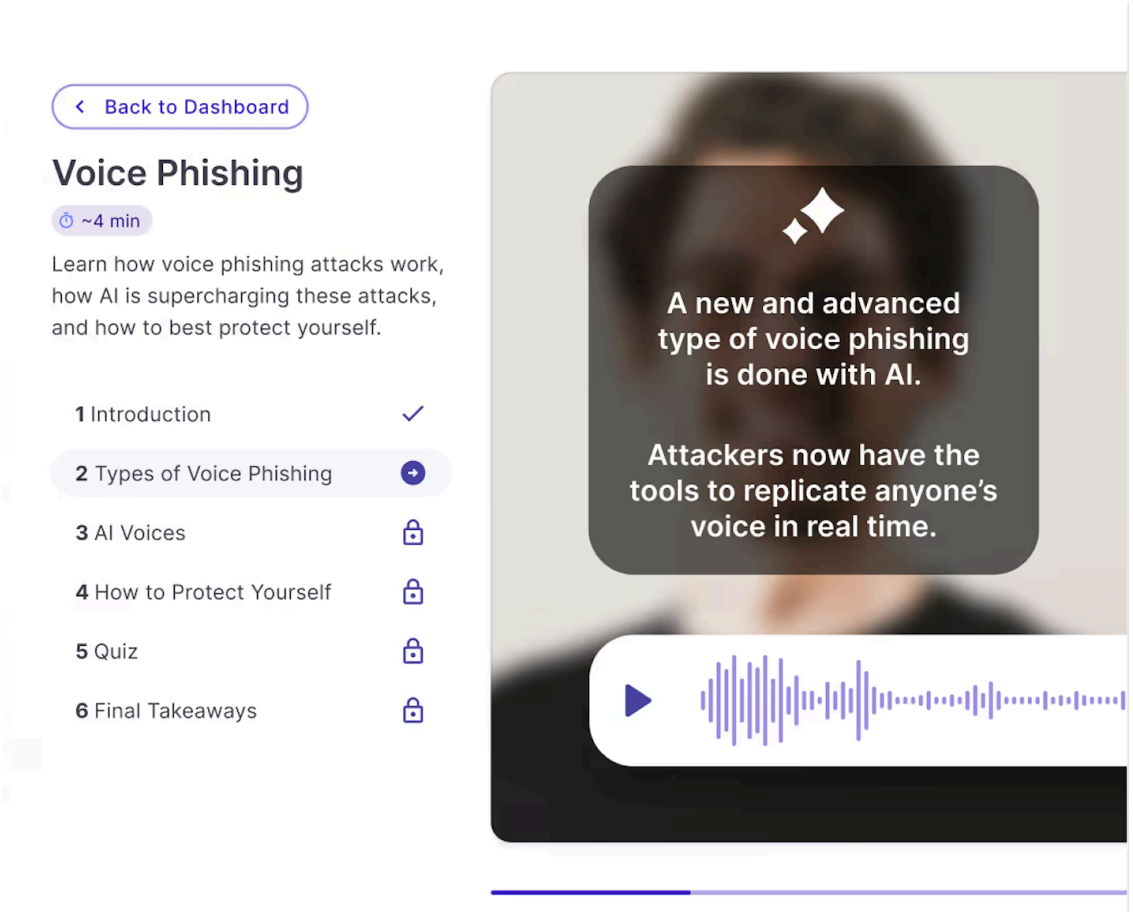
Why Adaptive Security is the Leading Defense Platform Against AI Phishing

Generative AI phishing scams are no longer clumsy attempts asking people to send money in exchange for a fake million-dollar payout. Today's generative AI scams involve multi-channel attacks delivered through email, phone, or even fake video calls.

That's where next-generation security awareness training tools from Adaptive Security help. Especially built for AI phishing, Adaptive simulates the tactics attackers now use, including deepfake voicemail requests and AI-crafted spear phishing emails.

Training scenarios are role-based and context-specific, so a finance team might see invoice fraud attempts while IT staff might test credential harvesting lures. This realism prepares employees to practice responding to them in conditions that feel real.

This is why forward-thinking organizations are already moving away from legacy platforms like KnowBe4 and Proofpoint. Instead, they're using Adaptive Security to give their team experience with new AI-generated cyber threats. The result is staff who don't freeze or fall for over-personalized details, but verify and respond correctly.



Adaptive Security AI voice phishing training module (Source)

Ready to see how your company can deal with evolving threats in real-time? Request a demo and experience how Adaptive prepares your teams for the phishing threats of the AI era.

Frequently Asked Questions: AI Phishing

How can I tell if a phishing email is AI-generated?

Look for overly polished, hyper-personalized details that feel unnecessary, like references to a recent conference talk or your exact job title. Tone that's too formal or phrasing that sounds rehearsed is another red flag.

What role do deepfakes play in phishing?

Deepfakes make phishing more convincing by exploiting trust in familiar voices and faces. Attackers can clone a CEO's voice to request a wire transfer or use synthetic video to impersonate leaders on a call. To avoid that, always confirm high-risk requests via a trusted second channel.

Is security awareness training actually effective against AI phishing attacks?

Yes, but only when it evolves with the cybersecurity threats. Modern platforms like Adaptive Security simulate AI-driven threats so employees practice responding under realistic conditions. That experience makes all the difference.

As experts in cybersecurity insights and AI threat analysis, the Adaptive Security Team is sharing its expertise with organizations.
