

Verifying Identity in the Age of **AI-Powered** Impersonation

A practical guide to today's AI-powered identity attacks and how to defend against them.



Specops, an Outpost24 company, is the leading provider of identity management and authentication solutions.

What's Inside?

Executive Summary

Chapter One: How AI Has Changed the Economics of Impersonation

Chapter Two: The Most Common AI-Powered Attacks Today

Chapter Three: Credentials Prove Only What You Know, Not Who You Are

Chapter Four: The High-Risk Moments Organizations Often Overlook

Chapter Five: A Guide to Stress-Testing Your Own Identity Verification Process

Chapter Six: How Specops Drives Resilience Against AI Identity Attacks



Executive Summary

Generative artificial intelligence (AI) is no longer a peripheral concern for security leaders. It is now a measurable feature of how attacks are built and run. [Verizon's 2026 Data Breach Investigations Report \(DBIR\)](#) found that the typical attacker now draws on AI assistance across roughly 15 distinct stages of an attack, with the most active users leaning on it for 40 or more. Working with threat-intelligence data from Anthropic, the same report identified phishing as the single largest use case among AI-assisted intrusion attempts, accounting for 44% of them, ahead of credential abuse and vulnerability exploitation combined.

For identity verification specifically, this shift has a very literal effect: seeing and hearing are no longer reliable proof of who someone is. Cloned voices, deepfake video, and synthetic or manipulated identity documents can now stand in for the sight and sound-based checks organizations have always leaned on. Compounding the problem is that the technology to produce convincing AI-driven attacks is cheap, fast, and within reach of most attackers. Yet many organizations are still defending this new threat with pre-AI methods: shared secrets, personal information, human judgment, and less secure multi-factor authentication (MFA) factors such as SMS codes and push notifications.

The AI-threat is accelerating in scale and sophistication. What once required a single, carefully maintained identity capable of surviving four rounds of interviews has become an operation that manufactures and discards identities on demand. On August 12, 2026, a [Wall Street Journal investigation](#) found a single North Korean hiring cell cycling through seven identities in one week and filing over 1,000 job applications in three months.

Addressing this problem starts at the small number of moments where identity is established or re-established: onboarding a new hire, issuing a first-day password, resetting a credential, recovering a locked account, and verifying a caller to the service desk. To help organizations develop a framework for meeting the challenge of verifying identity in an increasingly AI-driven threat landscape, this white paper examines the following:

- Ways AI-powered impersonation is reshaping identity risk, with the most common attack types and real-world examples
- Reasons traditional, credential-based verification methods are failing to keep pace
- High-risk moments in the employee identity lifecycle that organizations often overlook
- A framework for stress-testing an organization's own identity verification process against these risks
- Approaches organizations can take to build resilience against AI-driven identity attacks across the employee lifecycle



Chapter One: How AI Has Changed the Economics of Impersonation

Attackers Are Using AI to Simply Ask To Get Inside *AI Is Not Creating New Attacks, But It Is Accelerating Old Ones*

In November 2025, Anthropic reported that a state-sponsored group had manipulated its Claude Code tool to run an AI-orchestrated cyber espionage campaign against roughly 30 targets, with AI executing an estimated 80-90% of the operation. This is the kind of disclosure that fuels the idea that AI is creating an entirely new category of cyber threat. However, upon closer examination, the underlying techniques that enabled this attack (reconnaissance, credential harvesting, exfiltration) are not new.

Despite the fear, uncertainty, and doubt these AI developments generate, the data tells a more nuanced story. Verizon's 2026 DBIR found that the median AI-assisted technique had 55 known, pre-existing examples of tooling doing the same job, with fewer than 2.5% involving anything genuinely novel. AI is making familiar techniques faster, cheaper, and easier to scale, not inventing new ones.

That should not be read as an assertion that the threat landscape will stay this way indefinitely. In July 2026, Anthropic disclosed that its AI models had compromised real company infrastructure without any attacker directing them to at all, showing how quickly the ceiling on AI's offensive capability is rising. But for now, the practical implication for security leaders is that the gaps AI is currently exploiting are the same gaps that have gone unaddressed for years: assumed identity, static secrets, and human judgment calls made under pressure.

Sight and Sound Are No Longer Reliable Proof

For as long as security has existed, a recognized face or a recognized voice was close enough to proof. AI has broken that assumption, opening an identity gap between what organizations use to verify someone and what can now be convincingly faked. As AI rapidly collapses the skill and effort required to produce a fake voice, face, or message, organizations need to rethink their approach to identity, yet many still authenticate people exactly as they did a decade ago.

The scale of that gap is already showing up in the numbers. A 2025 Gartner survey of cybersecurity leaders found that 62% of organizations experienced a deepfake attack involving social engineering or exploitation of automated verification in the past 12 months. CrowdStrike's threat intelligence research found that number climbing fast: AI-enabled adversaries increased their attack volume by 89% year over year, while vishing-based intrusions doubled in the first half of 2026 alone.

In a Harvard-affiliated study, AI-automated spear-phishing emails achieved a 54% click-through rate, compared with 12% for generic phishing emails, matching the performance of human expert-crafted phishing at a fraction of the cost and skill required.



Chapter Two: The Most Common AI-Powered Attacks Today

Same Threats, New Delivery

Each of these four attacks follows the same shape: a plausible method turned into a real breach. What changes from one row to the next is only the delivery mechanism, voice, video, a fabricated identity, a well-timed message.

Attack	Method	Recent Example
Voice Cloning and Vishing	Attackers generate a synthetic voice from audio scraped off podcasts, keynotes, or voicemail greetings, then use it to impersonate executives, IT staff, or colleagues over the phone, often to push through an urgent, unverified request	Citadel, Point72, Millennium, and Two Sigma, August 5, 2026: attackers cloned executives' voices to try to extract credentials and network access in a coordinated campaign; Two Sigma blocked the attempt, and FINRA activated its new Financial Intelligence Fusion Center for the first time in response
Deepfake Video	Footage scraped from earnings calls, webinars, and press interviews becomes the raw material for a video call populated with fabricated, familiar-looking participants	Singapore, March 2025: \$499K wired after a Zoom call with a fabricated CFO and executives
AI-Enhanced Social Engineering	Researching a target in real time, attackers draft a tailored pretext and adapt their story as the conversation unfolds	Aura, March 2026: a vishing call gave ShinyHunters roughly an hour of access, exposing ~900K records
Synthetic and Stolen Identities	Stolen personal data, combined with fabricated or manipulated identity documents, gets attackers past hiring checks, video interviews, and background screening	US government agency, August 2026: the FBI disclosed that a North Korean worker had been hired as a remote IT contractor and operated inside a US government agency for several months before being discovered

What These AI Attacks Tell Us

- **The scale is growing:** Reaching dozens of targets once meant tailoring a pretext by hand for each one. AI has cut that cost, for example enabling a single campaign to hit Citadel, Point72, Millennium, and Two Sigma, along with several private equity firms, all at once.
- **Identities can be easily forged:** Sustaining a single identity through a background check and four interviews used to be the upper limit. [Verizon's 2026 DBIR](#) estimates fabricated and stolen identities have now been used across as many as 15,000 people to land remote roles at legitimate companies.
- **Detection is getting harder:** Spotting a fake used to be fairly easy. A [56-study meta-analysis](#) found humans now catch deepfake video only 57% of the time, barely above chance.
- **Entry points are expanding:** Email phishing was long the default way in. Google Cloud's [Mandiant M-Trends 2026](#) found voice phishing has overtaken it, now the second most common initial-access vector.

Deloitte's Center for Financial Services projects generative-AI-enabled fraud losses in the US will grow from \$12.3 billion in 2023 to \$40 billion by 2027. In each of these attacks, what failed was not identity itself, but the burden that was being placed upon it. The next chapter looks at exactly where that weight has been resting, and why it can no longer hold.

Chapter Three: Credentials Prove Only What You Know, Not Who You Are

Traditional Verification Methods Were Not Designed for Today's AI-Dominated Threat Landscape.

What Used to Be Secret Is Now Public Record

Most identity controls in the enterprise are built to authenticate, meaning they check that someone possesses the right password, the right token, or the right one-time code. Authentication answers the question of whether a person has the correct credential. It does not answer whether the person presenting that credential, or asking to be issued one, is actually who they claim to be.

That second question, identity verification, has historically been treated as a one-time event, such as a new starter's initial password setup. In practice, organizations are required to re-establish trust in someone's identity far more often than that:

- When a new hire needs their first credentials
- When a password or MFA method is reset
- When an account is recovered
- When a caller reaches out to the service desk claiming to be a colleague

Many of these moments still rely on knowledge-based verification, such as a name, a date of birth, an employee ID, or a manager's name. The data reflects how exposed that approach has become. Verizon's 2026 DBIR found that 73% of ransomware victims had an associated infostealer infection or credential leak in the year before the attack, with half of those occurring within 95 days of the ransomware event itself.

The [Identity Theft Resource Center](#) separately tracked 1,803 data compromises in the first half of 2026 alone, putting the year on pace to exceed 2025's record of roughly 3,321 compromises. Each compromise adds to a pool of personal information that attackers, and the AI tools now doing their research for them, can draw on to answer a "secret" question convincingly. What used to function as a secret is now, in practical terms, public record.

People Remain the Weakest Link in Security

Individuals are being asked to make judgment calls in real time that most people would struggle with under any circumstances. Advanced deepfakes, AI-assisted scripts built around genuine personal details, and well-rehearsed threat actors make it extremely difficult for service desk agents to reliably distinguish a genuine caller from an impostor.

In June 2025, a vishing call to a third-party contact center handling Qantas customer service persuaded an operator to grant account access, exposing the records of roughly 5.7 million customers. The attackers, widely attributed to the Scattered Spider collective,



had already gathered the personal details they used from public sources and prior breaches.

None of this means passwords, MFA, or human service desk agents are obsolete. They remain essential layers of enterprise security. It means that the moments when normal authentication is unavailable, not yet established, or being reset deserve their own, stronger form of verification, because they are precisely the moments an attacker can use to work around otherwise strong controls.

Chapter Four: The High-Risk Moments Organizations Often Overlook

Attackers Are Excellent at Playing the Waiting Game

New-Hire Onboarding

Every step of onboarding happens for the first time, often under time pressure, with someone the organization may never have verified in person. There is no established pattern of behavior to compare a request against, and no history that would make an unusual request stand out.

A new hire is, by definition, unknown to IT and unfamiliar with company process, and that combination, a new relationship, urgency, and no baseline, is precisely what identity attacks are built to exploit.

AI exploits this specifically by manufacturing the missing baseline itself: a synthetic photo, a fabricated document, or a cloned voice convincing enough that no one has reason to question it.



Service Desk Password and MFA Resets

A service desk reset does not only restore access to one account. It typically restores whatever that account was already entitled to, which for many employees includes broad or privileged reach across the network. AI exploits this by giving attackers a convincing cloned voice and a well-researched, personalized story to work from in real time, on the call itself, exactly the pressure an agent has only seconds to evaluate before deciding whether to act.

Why Wait?

What makes these moments worth waiting for, rather than searching for a technical exploit, is that they do not depend on a flaw existing. A vulnerability might not be there, or might already be patched. But every organization runs a hiring process, issues credentials, resets passwords, and staffs a service desk, which means these moments are not a possibility an attacker has to create, they are certainty an attacker only has to wait for.

AI has shortened that wait and widened the payoff: it can manufacture the missing baseline at onboarding, produce a convincing voice at the service desk, and do both cheaply enough to try again and again until one attempt lands.

Chapter Five: A Guide to Stress-Testing Your Own Identity Verification Process

Each row in this framework reflects an identity interaction that has already been exploited at a real organization.

The pattern across all six interactions is the same: verification relies on a credential, a script, or a static fact that AI has made straightforward to imitate or bypass. To stress-test your own process, work through each row and compare it against your organization's current approach.

Any row where your process matches, or comes close to, the typical verification approach in the middle column represents an active exposure, and the recommended control shows what closes it.

Identity Interaction	Typical Verification Approach	Impersonation Risk	Recommended Control
New-hire onboarding	Manager vouching, video interview, document upload	Deepfaked photo/video, stolen or synthetic identity documents	Government-ID scanning + biometric liveness before access is granted
First Active Directory password	IT creates password, shares via email, SMS, or call	Interception of the shared credential; SIM-swap; phishing	New hire sets their own first password after identity is verified; IT never shares the password
Password reset	Agent asks name, employee ID, date of birth	AI-scraped personal data and cloned voice used to pass caller checks	Technical verification gate required before any reset action is possible, using self-service, security-focused MFA that works remotely from any device
MFA / device reset	Manual agent judgment over the phone	Single point of bypass for otherwise strong MFA controls	Same verification gate, plus out-of-band confirmation for privileged accounts
Account recovery	Static security questions, personal information	AI can research and answer many "secret" questions	Self-service recovery that do not rely on static personal facts
Service desk request	Caller states name/ID, answers questions	Voice cloning and vishing impersonating a known employee	Seamless verification required before any sensitive action, regardless of urgency

Prioritize By Risk, Not By Convenience

Not every identity interaction needs the same standard of verification, and applying government-ID scanning and biometric liveness everywhere adds friction without a proportionate security benefit. The six moments in the framework above are not a checklist to implement all at once. They are a starting point for prioritizing rollout by exposure.

The clearest place to start is wherever a single credential handoff or a single human judgment call currently creates the biggest point of failure across an organization. This is typically new-hire onboarding and service desk resets, both of which this paper has shown can hand an attacker broad, often privileged, access from one successful impersonation.

Verifying a new hire's identity once, at the start of employment, is a reasonable one-time cost. Applying the same bar to every routine password reset would create disproportionate friction for what is usually a low-risk request. Biometric liveness detection should be understood for what it is: confirmation that a live human is present, not a general-purpose deepfake detector. No single control in the framework closes every gap on its own, which is why a layered approach across all six moments matters more than perfecting any one of them in isolation.

Chapter Six: How Specops Drives Resilience Against AI Identity Attacks

Specops Replaces Assumed Identity Across The Entire Employee Identity Lifecycle.



Verifying Identity Before Access Begins

Specops Secure Onboarding, powered by Specops Verified ID, lets new hires set their own first Active Directory password, and only after their identity has been verified. IT never creates or shares a password on their behalf, removing the interception risk that email and SMS handoffs carry. Identity is confirmed through government-ID scanning and biometric liveness detection, covering more than 16,000 document types across 254 countries and territories, and is cross-checked against the user's Active Directory or Entra ID record.

This is not a replacement for an organization's broader identity and access management (IAM) or authentication stack. It is purpose-built for this specific high-risk moment, and integrates with an organization's existing IAM stack to close the gap at that moment rather than duplicate what is already in place. The underlying AI and machine learning models that power this solution are pre-trained and do not use or retain customer data, addressing a common trust concern associated with biometric verification.

Remove Human Judgment Call at the Service Desk

Specops Secure Service Desk technically blocks agents from taking action, such as a password or MFA reset, until caller identity is confirmed, removing the guesswork this paper has shown attackers rely on. It integrates with Okta, PingID, Duo, and Symantec VIP for verification, alongside ServiceNow and Jira workflows, so identity checks sit inside the ticketing process agents already use rather than adding a separate step.

The same verification events that build an audit trail can also be sent to SIEM, SOC, and

analytics platforms through a dedicated Security Event API, giving security teams visibility into verification attempts and failures rather than relying on agents to flag suspicious calls after the fact.

With multi-language support and a customizable interface, the control holds consistently across a global service desk, rather than depending on the judgment, script, or vigilance of whichever individual agent happens to take the call.

Make Recovery Self-Service, Not Guesswork

Specops uReset replaces the static personal facts discussed throughout this paper with a flexible MFA engine supporting more than 20 authentication factors, including third-party identity providers such as Okta, Duo, PingID, and Symantec VIP. Users can reset a password or unlock an account from any device, browser, or location, on or off VPN. Users can reset a password or unlock an account from any device, browser, or location, on or off VPN. For users without a personal device to authenticate with, uReset offers alternatives such as trusted network location and manager identification, rather than falling back on the security questions this paper has shown AI can now answer at scale. Where security questions remain necessary, it adds protections such as withholding all questions at once and obfuscating typed answers.

Built-in Breached Password Protection checks new passwords against a database of more than 6 billion known compromised credentials, so a self-service reset cannot simply reintroduce a password already exposed elsewhere. Dynamic, real-time policy feedback then helps users choose a compliant password on the first attempt, rather than escalating a failed reset into a service desk call. The platform works across on-premises Active Directory, hybrid, and cloud-only Entra ID environments, and integrates with Secure Service Desk so that recovery requests needing a human, rather than self-service, still pass through the same identity verification.

This matters operationally as much as it does securely: Gartner estimates that up to 40% of all IT service desk calls are password related, so removing the guesswork from this moment reduces attacker opportunity and service desk load at the same time.



Get Started With Specops

Specops Secure Onboarding, Secure Service Desk, and uReset are not three separate tools addressing three separate problems. They close the same gap at three different moments in the employee identity lifecycle: an assumed identity at hiring, a static secret at reset, and an unverified caller at the service desk.

Handled individually, each of these remains a point an attacker can work around. Handled together, they remove the assumption of trust from every moment identity is established, reset, or recovered, rather than defending it at just one.

Organizations looking to extend this discipline further can apply the same logic to the layers surrounding identity itself, from the strength of the credential to the trustworthiness of the device requesting access. Specops Device Trust addresses this latter layer directly, binding verified identities to authorized devices and continuously checking device posture throughout a session, rather than only at login. But the starting point remains the same: replace assumed identity with verified identity at the moments that matter most.

Find out how Specops can help your organization verify identity at the moments that matter most.

[Book a Demo](#)[Speak to an Expert](#)

Specops

Specops Software, an Outpost24 company, is the leading provider of identity management and authentication solutions. Specops protects your business data by blocking weak passwords and securing user authentication. With a complete portfolio of solutions natively integrated with Active Directory and Entra ID, Specops ensures that sensitive data is stored securely, either on-premises or within your Entra ID tenant, and remains under your control. Specops Software was founded in 2001 and is headquartered in Stockholm, Sweden with additional offices in the US, Canada, the UK, and Germany.