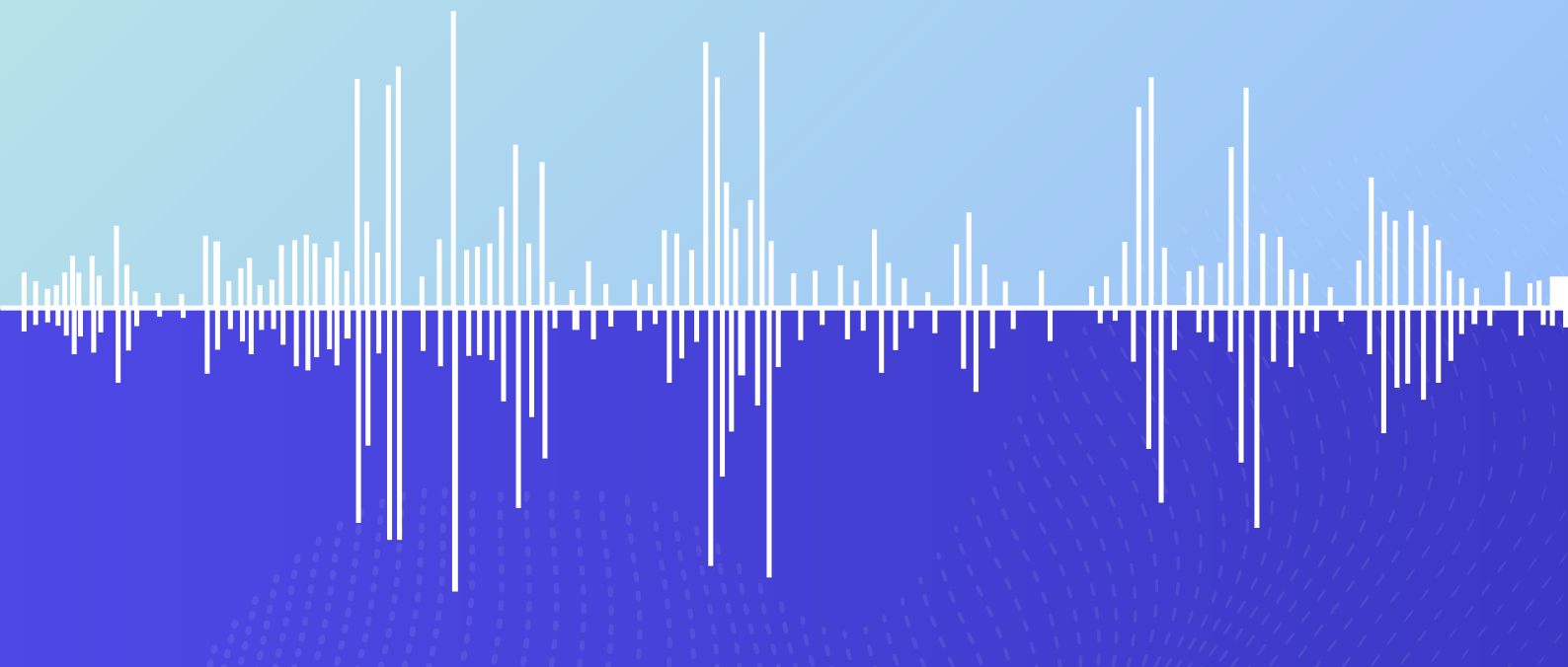
MIRAGE

AI SOCIAL ENGINEERING STRATEGY GUIDE



INTRODUCTION

AI social engineering isn't science fiction. It's already here, and it's being used against your workforce.

This guide isn't about adding another layer of surveillance or creating fear. It's about helping your organization safely experience the future of social engineering so you can respond with insight, not instinct.
Done right, these simulations don't erode trust. They reinforce it.

They don't punish mistakes. They reveal opportunities.

Like any good learning experience, rhythm matters. Don't wait for an incident. Sequence micro-simulations, nudges, or moments of exposure over time. Familiarity builds fluency.

Start simple. Start with empathy.
And start before someone else does it for real.

— Ross Lazerowitz

PROGRAM MISSION

The goal of this program is to introduce AI-driven social engineering simulations in a way that builds trust, not fear, across the organization.

These simulations are not designed to “catch” users or trick teams. They’re a safe, ethical way to explore how modern threats are evolving, how humans respond under pressure, and how we can collectively improve.

By focusing on education over punishment, realism over theatrics, and transparency over deception, this program helps:

- Security teams validate detection and response processes
- HRM and Awareness leaders give users a chance to experience modern threats without real-world risk
- Red teams run accurate tabletop exercises grounded in today’s attacker tactics

At its core, this program is about helping people feel more prepared, not more paranoid.

Program Ownership Note

Ownership varies. In organizations with active Red Teams, simulations may live under offensive security. In others, HRM or Awareness leads may own the strategy and delivery. What matters most is having someone who understands both human behavior and simulation design and is empowered to build trust, not just test users.

WHAT IS AI SOCIAL ENGINEERING?

AI social engineering refers to the use of generative technology, such as AI-generated voice, language models, or synthetic identities, to simulate real-world attacker behavior in a safe, controlled environment.

These simulations go beyond email. They replicate modern threat techniques, including phone calls, text messages, social media engagement, and even live dialogue. The goal isn’t to surprise users, but to mirror the methods that criminals are now using so that teams can observe, respond, and improve.

AI social engineering can be:

- Pre-recorded or real-time
- Fully autonomous or partially guided
- Used for employee awareness or risk assessment

It is fundamentally about preparedness, giving people and systems a chance to experience the future of social engineering in a way that drives insight, not fear.

This isn’t phishing 2.0. It’s a new class of simulation built around conversation, context, and credibility, not clickbait.

ETHICAL FOUNDATIONS

AI social engineering carries real power. That power needs structure, not just in how it's used, but in how it's perceived.

This program is built on three core ethical pillars:

Transparency of Intent

Employees should understand the purpose of these simulations. Communicate that:

- AI-based scenarios are part of the awareness or testing program
- The goal is education, not entrapment
- Participation is monitored only to improve processes and preparedness

Respectful Simulation Design

Simulations should never humiliate or corner users. Stick to plausible scenarios rooted in work reality. The goal is to create teachable moments, not emotional discomfort.

Role-Based Consent

Not every employee needs to experience every simulation. Use role-based scoping and where warranted, provide document opt-in or opt-out options.

Expert Guidance
Look for vendors that offer built-in opt-out or experience-alternative pathways such as video walkthroughs or guided debriefs.

What If the Risk Isn't Malicious? Designing for Humanity

If a user struggles to respond or shows signs of distress, it may not be a training failure. It could be a signal. When simulation outcomes suggest someone is overwhelmed, confused, or emotionally depleted, that can be a moment for HR or People Ops to step in with support. Design your program with the assumption that sometimes the biggest risk isn't a malicious actor, it's an employee quietly struggling.

TWO PATHS, ONE PURPOSE

AI social engineering simulations serve two strategic purposes, depending on who's driving the program.

- **Red Team Led** → Assessment Focus
  These teams deploy AI simulations to emulate real-world adversaries, test technical controls, and validate response paths. The outcome is diagnostic. The experience is secondary.
- **HRM or Awareness Led** → Experience Focus
  These teams prioritize safe exposure. The goal is to let users feel what it's like to face a believable attacker and walk away more confident, not more compliant.

Both are valid. Both can co-exist.
The key is defining which path you're on or when it makes sense to switch.

USE CASES BY PERSONA

For Red Teams: Simulated Adversary Engagement

AI social engineering allows red teams to move beyond phishing templates. These tools simulate how real attackers operate through voice, SMS, or dynamic, real-time pretexting.

Use cases:

- Simulate reconnaissance, vishing, or behavioral probing
- Observe user instincts under pressure
- Test escalation paths, tooling, and business processes.

Some red teams may align lightly with HRM or awareness, which leads to shaping scenarios or sharing observations.

For HRM and Awareness Leaders: Safe Exposure to Real Threats

These simulations give users a chance to experience modern threats without shame, punishment, or damage.

Benefits:

- Enable experiential learning grounded in realism
- Build awareness of manipulation cues
- Encourage behavior change through exposure, not fear

Some HRM teams choose to collaborate with red teams for scenario design or review. This isn't mandatory, but it can improve cultural alignment.

For CISOs and Executive Leaders: Validating Readiness Without Compromising Trust

AI social engineering simulations allow leaders to:

- Validate detection and response across human and technical systems
- Reveal blind spots in workflows or culture
- Demonstrate a proactive, empathy-forward security program
- And most importantly, provide actionable metrics

When designed and delivered well, these simulations can enhance trust in leadership, not erode it.

NOTE

AI social engineering is not a replacement for phishing simulation. It is the next step. Most phishing programs have plateaued. This evolution allows you to create targeted, behavior-driven experiences that actually reflect today's attacker techniques.
Where phishing simulations were spray-and-pray, this is surgical. It is a strategic opportunity to explore why users act the way they do, not just if they clicked.

Vendor or Internal Execution

Determine whether simulations are better run by internal red teams or vendors. Look for partners who:

- Respect privacy and PII
- Support role-based targeting
- Provide scrubbed, client-isolated logs

PROGRAM DESIGN AND DELIVERY

Define the Objective First

Examples:

- Test recognition of voice-based phishing
- Observe escalation and reporting
- Provide users with safe exposure to attacker tactics

Tone and Experience Guidelines

- Keep it casual, grounded, and internal-feeling
- Avoid dramatic, manipulative, or fear-based pretexts
- Prioritize scenarios that could actually happen to that user type

Audience and Role-Based Scoping

Examples:

- Executive assistants → spoofed exec call
- Developers → impersonated IT SMS
- Finance → voice pretext around payment verification

Scenario Duration

Suggested runtime: 2 to 5 minutes

- Enough time for the user to engage and react
- Avoid emotional exhaustion or fatigue

User Notification Strategy

- **Before:** Org-wide announcement of upcoming AI simulation program
- **During:** Simulation occurs with opt-out paths and internal safety nets
- **After:** Follow-up with guidance, praise, and behavioral reinforcement

Delivery Format

- Voice, SMS, chatbot, or hybrid
- Manual execution or platform-driven
- Consider Slack or Teams for nudges or follow-ups in users' natural workflows

MEASURING SUCCESS

Human-Centric Metrics

- Recognition rate
- Reporting path accuracy
- Escalation quality
- Post-event confidence and reflection

Technical and Process Metrics

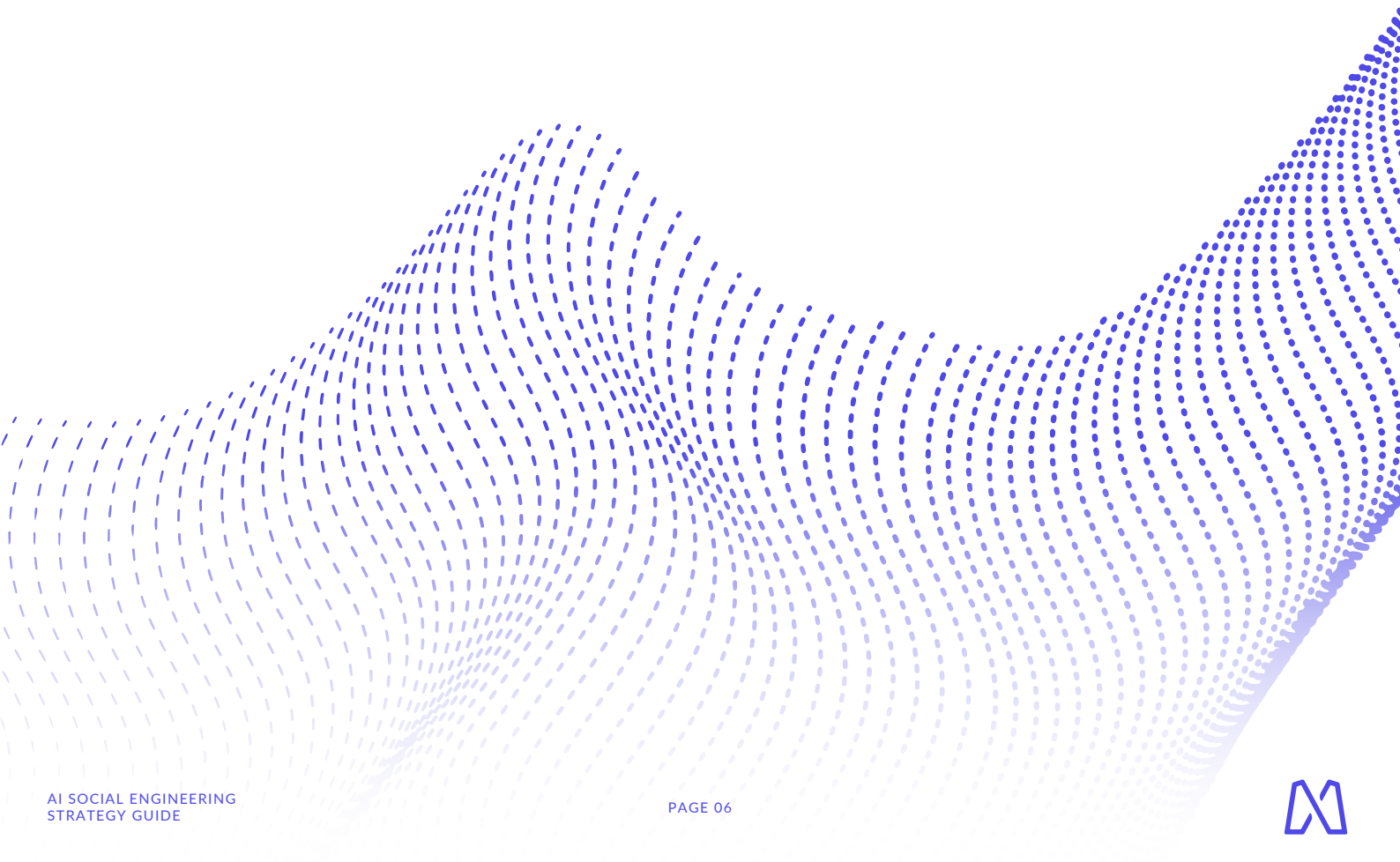
- Detection time by SOC or help desk
- Tooling effectiveness
- Escalation chain performance
- Interdepartmental handoffs

Readiness Indicators

- Users discuss the event with peers or family
- Culture of reporting improves
- Feedback reflects increased confidence

Wellness Signal (Optional)

Unusual responses may indicate emotional fatigue or deeper issues. If surfaced ethically and supportively, this insight can guide HR toward compassionate outreach.



BRIDGING DISCIPLINES

AI simulations work best when they're not siloed.

Security and HRM

- Align tone and risk goals
- Support user empowerment, not punishment

Red Teams and Awareness

- Combine realism with learning
- Awareness owns the narrative, red team owns the realism

Legal, People Ops, Comms

- Set ethical boundaries
- Help craft internal messaging
- Prevent surprise or unintended consequences

STAKEHOLDER ALIGNMENT CHECKLIST

Security or Red Team

- Define scope and goals
- Approve tools and oversight

HRM or Awareness

- Shape tone and support options
- Confirm opt-out design

People Ops or HR

- Address emotional safety
- Set boundaries

Legal or Compliance

- Data privacy review
- Consent handling

Corporate Comms

- Internal messaging
- Post-simulation guidance

IT or Help Desk (Optional)

- Escalation monitoring
- Internal support alert

TIPS FROM OUR CO-FOUNDER AND CEO

We've worked with organizations ranging from startups to hundred-thousand-person public companies and run thousands of simulations. In the process, we've learned a thing or two. The following are our best practices for running simulations and what to do afterward.

TIP #1

Step into the attacker's shoes

As a security professional, you have a much harder job than the attackers. You have to operate within your organization's guidelines, deal with budgets and approval, and think about organization-wide risk. When scoping out an AI Social Engineering simulation, it's helpful to put all that aside and look at your risks like an attacker would. They aren't trying to run a monthly phishing assessment, they are after something - money, information, reputation harm, etc. Think about your organization's "nightmare" scenarios and enlist the help of other departments to threat model. Prioritize and stack rank these scenarios instead of going organization-wide on day one.

TIP #2

Smoketest your processes

Training is essential, but it's never a full substitute for strong business and technical controls. Use simulations as smoke tests for your organization's actual response pathways: Are reporting channels intuitive? Can the help desk quickly escalate? Does your security team know how to respond when a realistic threat emerges? Often, you'll uncover gaps in communication, tooling, or even the cultural expectations around reporting. Make sure your program focuses as much on validating the back-end workflows and tooling as it does on user awareness.

TIP #3

Stop assuming users are the weakest link

Security teams often default to seeing employees as vulnerabilities, but in reality, they can become your greatest assets. When running simulations, prioritize rewarding good behaviors over punishing mistakes. Publicly celebrate accurate identifications, timely escalations, or even thoughtful questioning. Recognizing employees who demonstrate proactive security behaviors creates a culture of positive reinforcement, turning your workforce into active participants rather than passive targets. Over time, you'll cultivate security champions who naturally reinforce strong practices across your organization.

— Ross Lazerowitz

MIRAGE

TALK TO OUR FOUNDERS

ASK US ANYTHING

+1 707-563-9264
hello@miragesecurity.ai
miragesecurity.ai