

Social Engineering & Phishing with AI

How AI supercharges the oldest attack vector in cybersecurity — and how defenders can keep up.

Educational Use Only

Social engineering and phishing techniques are demonstrated for authorized red team engagements and CTF competitions ONLY. Phishing real individuals without explicit authorization is illegal and causes real harm. All examples use fictional targets.

Red Team

Social Engineering

Phishing

AI-Assisted Attacks

OSINT

Notes: Emphasize the legal and ethical boundaries from the very start. Every technique shown requires explicit written authorization.

Agenda

1. **Social engineering principles** — the psychology behind the attack
 2. **AI-generated phishing emails** — LLMs as pretext engines
 3. **Spear phishing with OSINT** — personalizing attacks at scale
 4. **Vishing & smishing** — voice cloning and SMS-based attacks
 5. **Phishing infrastructure** — domains, landing pages, credential harvesting
 6. **AI-powered pretexting** — building believable personas
 7. **Realistic examples** — templates, HTML snippets, campaign anatomy
 8. **Detection & defense** — understanding both sides
 9. **Ethics & legal boundaries** — rules of engagement
 10. **CTF challenge** — craft a targeted phishing campaign
-

Why Social Engineering Still Works

Technical defenses have improved dramatically. Humans have not.

- **91% of cyberattacks** begin with a phishing email (Deloitte, 2023)
- **Social engineering bypasses firewalls** — it targets the human layer
- **AI lowers the barrier** — fluent English, perfect grammar, cultural awareness
- **Scale + personalization** — what once took hours per target now takes seconds

The Fundamental Truth

You don't hack the system. You hack the person who has access to the system.

Cialdini's Principles of Influence

Robert Cialdini identified six principles that drive human compliance. Social engineers weaponize all of them.

Principle	Description	Phishing Example
Authority	We obey figures of authority	"CEO requests urgent wire transfer"
Urgency / Scarcity	Fear of missing out or consequences	"Account suspended — verify in 24h"
Reciprocity	We feel obligated to return favors	"Here's the report you asked for" (with payload)
Social Proof	We follow what others are doing	"Your colleagues have already updated their info"
Liking	We comply with people we like	Friendly, personalized tone from a "coworker"
Commitment	We honor prior agreements	"As discussed on our call last Tuesday..."

Notes: Have students identify which principles are at play in phishing samples they encounter in the wild.

Weaponizing Authority

Authority is the most powerful lever in business email compromise (BEC). Targets rarely question a message that appears to come from leadership.

Attack Pattern

- Impersonate C-suite executives
- Use display-name spoofing or lookalike domains
- Request urgent, confidential actions
- Discourage verification ("Keep this between us")

AI Enhancement

- LLMs mimic executive writing style from public speeches and posts
- Generate context-aware requests referencing real projects
- Produce flawless grammar in any language
- Adapt tone to match corporate culture

Urgency, Reciprocity & Social Proof

Urgency / Scarcity

- Creates panic that bypasses rational thinking
- "Your password expires in 1 hour" / "Unusual login detected"
- AI generates time-sensitive scenarios tailored to the target's role

Reciprocity

- Offer something valuable before making a request
- "I found this security report relevant to your team" (malicious attachment)
- AI creates convincing "gifts" — fake reports, resources, or introductions

Social Proof

- Reference colleagues, departments, or industry peers
- "Finance already submitted theirs — you're the last one"

- AI scrapes org charts to name real team members
-

AI-Generated Phishing: The Paradigm Shift

Before LLMs, phishing had telltale signs: broken grammar, generic greetings, implausible scenarios. AI eliminates all of them.

Traditional Phishing

- Template-based, generic
- Grammar and spelling errors
- Bulk spray-and-pray
- Easy to detect with filters
- Low conversion rate

AI-Enhanced Phishing

- Contextually personalized
- Perfect grammar in any language
- Scalable spear phishing
- Evades content-based filters
- Dramatically higher conversion

Notes: A 2023 study showed AI-generated phishing emails had a 3x higher click-through rate than human-written ones.

How LLMs Create Convincing Pretexts

An attacker provides context; the LLM generates a polished, believable message. The workflow is simple.

1. **Gather OSINT** — target's name, role, company, recent events
2. **Define the pretext** — scenario, desired action, urgency trigger
3. **Prompt the LLM** — request a professional email in the right tone
4. **Review and weaponize** — add malicious links or attachments
5. **Send via spoofed infrastructure** — matching domain and headers

```
# Example: Generating a pretext with an LLM (educational demonstration)
prompt = """
Write a professional email from Sarah Chen, VP of Engineering at Nexus Corp,
to a new developer named Alex Rivera. The email should:
- Welcome them to the team
- Ask them to complete mandatory security training
- Include a link to the 'training portal' (use placeholder URL)
- Sound warm but corporate
"""
```

The LLM produces a perfectly crafted, contextually appropriate email
that is nearly impossible to distinguish from legitimate corporate mail

Example: AI-Generated Phishing Email

Below is a fictional example of what an AI-assisted phishing email might look like targeting a new employee at a fictional company.

From: Sarah Chen <s.chen@nexus-corp.com>
To: Alex Rivera <a.rivera@nexus-corp.com>
Subject: Welcome aboard! Action required: Security training

Hi Alex,

Welcome to the Nexus Corp engineering team! We're excited to have you on board.

As part of onboarding, all new engineers need to complete our mandatory security awareness training before EOD Friday. It only takes about 20 minutes.

Please sign in with your corporate credentials here:

<https://nexus-corp-training.com/onboard> [MALICIOUS - lookalike domain]

Let me know if you run into any issues. Looking forward to our 1:1 on Thursday!

Best,

Sarah Chen

VP of Engineering | Nexus Corp

Principles used: Authority (VP), urgency (EOD Friday), commitment (1:1 reference), liking (warm tone).

Spear Phishing with OSINT

OSINT transforms generic phishing into surgical spear phishing. AI accelerates the reconnaissance-to-attack pipeline.

OSINT Sources for Spear Phishing

Source	Intelligence Gathered	Attack Use
LinkedIn	Job titles, reporting chain, tenure	Impersonate manager, reference role
Company blog	Projects, launches, team names	Reference real initiatives in pretext
GitHub/GitLab	Tech stack, commit history, collaborators	Craft developer-specific lures
Social media	Interests, travel, life events	Build rapport, personalize tone
SEC filings / press	M&A activity, leadership changes	BEC pretexts around transitions

Source	Intelligence Gathered	Attack Use
Job postings	Tools used, team structure, pain points	Pose as vendor for listed tools

OSINT-to-Phishing Pipeline

AI automates what used to be a manual, hours-long process for each target.

```
# Fictional demonstration of an OSINT-to-phishing pipeline
target_profile = {
    "name": "Jordan Blake",
    "role": "Senior DevOps Engineer",
    "company": "Fictional Acme Systems",
    "manager": "Pat Morgan (Director of Infrastructure)",
    "tech_stack": ["Kubernetes", "Terraform", "AWS"],
    "recent_activity": "Presented at KubeCon 2025 on cluster autoscaling",
    "linkedin_connections": ["DevOps Weekly newsletter", "HashiCorp community"]
}

# AI generates a personalized phishing email
pretext = f"""
From: Pat Morgan <p.morgan@acme-sys.com>
Subject: KubeCon follow-up — Terraform module review
```

Every detail — the conference, the tech stack, the manager's name — came from publicly available information.

Vishing: AI Voice Cloning Attacks

Voice phishing (vishing) has been transformed by AI voice cloning. A few seconds of audio can now generate a convincing voice replica.

How AI Voice Cloning Works

1. **Sample collection** — 3-10 seconds of target's voice (earnings calls, YouTube, podcasts)
2. **Voice model training** — tools like ElevenLabs, Tortoise TTS clone the voice
3. **Script generation** — LLM writes a natural-sounding phone script
4. **Real-time synthesis** — attacker speaks; AI converts to cloned voice live

Real-World Impact

In 2024, a finance worker in Hong Kong transferred \$25 million after a video call with deepfake versions of the company's CFO and other executives. The entire call was AI-generated.

Notes: Voice cloning tools are commercially available. The barrier is awareness, not technology.

Smishing: AI-Powered SMS Attacks

SMS phishing (smishing) exploits the inherent trust people place in text messages and the limited UI of mobile devices.

Why Smishing is Effective

- **98% open rate** — SMS messages are almost always read
- **No hover-to-preview** — mobile users cannot inspect URLs easily
- **Urgency is native** — texts feel immediate and personal
- **Short format** — less room for red flags to show

AI-Enhanced Smishing Examples

[Acme IT] Jordan, your VPN certificate expires today. Renew now to avoid losing access: <https://acme-vpn-renew.co/cert> *[malicious]*

[Package Update] Your delivery from the AWS re:Invent merch store needs address confirmation:
<https://shipping-confirm.co/track> *[malicious]*

AI tailors SMS content based on OSINT — referencing real tools, events, or services the target uses.

Phishing Infrastructure: Domains

Convincing phishing requires infrastructure that looks legitimate. Domain selection is the first critical step.

Domain Spoofing Techniques

Technique	Example	Detection Difficulty
Typosquatting	nexus-c0rp.com (zero for 'o')	Medium
Homograph attack	nexus-corp.com using Cyrillic 'o'	High
Subdomain abuse	nexus-corp.login.attacker.com	Low-Medium
Lookalike TLD	nexus-corp.co (instead of .com)	Medium
Compromised domain	Hijacked legitimate domain	Very High
Cloud hosting	nexus-corp-sso.azurewebsites.net	High

```
# Tools for generating lookalike domains (authorized testing only)
```

```
# dnstwist – finds permutations of a target domain
```

```
dnstwist --registered nexus-corp.com
```

```
# urlcrazy – generates typo variations
```

```
urlcrazy nexus-corp.com
```

Phishing Landing Pages

The landing page must be a pixel-perfect replica of the legitimate login portal. AI and modern tools make this trivial.

Credential Harvesting Page (Simplified Example)

```
<!-- FICTIONAL EXAMPLE — for educational/CTF use only -->
<!DOCTYPE html>
<html>
<head>
  <title>Nexus Corp - SSO Login</title>
  <style>
    body { font-family: Arial; background: #f5f5f5; display: flex;
           justify-content: center; align-items: center; height: 100vh; }
    .login-box { background: #fff; padding: 40px; border-radius: 8px;
                 box-shadow: 0 2px 10px rgba(0,0,0,0.1); width: 380px; }
    .logo { text-align: center; margin-bottom: 24px; }
    input { width: 100%; padding: 12px; margin: 8px 0; border: 1px solid #ddd;
            border-radius: 4px; }
    button { width: 100%; padding: 12px; background: #0066cc; color: #fff;
            border: none; border-radius: 4px; cursor: pointer; }
```

Notes: Tools like Gophish and Evilginx2 automate the full credential harvesting workflow including MFA bypass via reverse proxying.

Credential Harvesting & MFA Bypass

Modern phishing does not stop at passwords. Sophisticated attacks capture session tokens to bypass MFA entirely.

Phishing Frameworks for Red Teams

Tool	Capability	MFA Bypass
GoPhish	Campaign management, email templates, landing pages	No
Evilginx2	Reverse proxy, real-time credential + token capture	Yes
Modlishka	Transparent reverse proxy phishing	Yes
EvilnoVNC	Browser-in-the-browser via VNC streaming	Yes

Evilginx2 Attack Flow

1. Victim clicks phishing link pointing to your reverse proxy

2. Proxy relays traffic to real login page — victim sees the real site
 3. Victim enters credentials + MFA code on the real site (through proxy)
 4. Proxy captures the authenticated session cookie
 5. Attacker imports cookie into their browser — fully authenticated
-

AI-Powered Pretexting

Pretexting is the art of creating a fabricated scenario to engage a target. AI makes pretexts richer, more consistent, and harder to detect.

Building a Believable Persona

- **Backstory generation** — LLMs create detailed, consistent personal histories
- **Profile photos** — AI-generated faces (StyleGAN, DALL-E) for social media
- **Digital footprint** — LinkedIn profile, GitHub activity, blog posts
- **Communication style** — AI maintains consistent tone across interactions
- **Domain expertise** — LLMs simulate deep knowledge of any field

```
# Fictional persona generation for authorized red team exercise
persona = {
    "name": "Morgan Ellis",
    "role": "Cloud Security Consultant",
    "company": "Sentinel Advisory Group", # fictional
    "linkedin_headline": "Helping enterprises secure their cloud journey | AWS | Azure",
    "backstory": "10 years in cloud infra, former SRE at [fictional] TechCo",
    "approach": "Connect on LinkedIn, build rapport over 2 weeks, then send
```

```
        'shared resource' with tracking payload"
    }
    # AI generates all communication – connection requests, messages, follow-ups
    # Each message maintains the persona's voice and builds trust incrementally
```

Multi-Channel Pretexting

The most effective social engineering attacks use multiple channels to build legitimacy and create a sense of familiarity.

Attack Chain Example

Day 1: LinkedIn connection request from "Morgan Ellis" — references shared KubeCon interest

Day 3: LinkedIn message — comments on target's recent post, shares a relevant article

Day 7: LinkedIn message — mentions an upcoming webinar on Kubernetes security

Day 10: Email — "Great connecting on LinkedIn! Here's the white paper I mentioned" (tracking

link)

Day 12: Follow-up email — "Did you get a chance to read it? I'd love your feedback" (credential harvesting link)

Each touchpoint reinforces the persona and increases trust. AI manages all channels simultaneously and maintains consistency.

Phishing Email Templates: Common Pretexts

These are the most effective pretext categories used in real-world red team engagements.

Pretext Category	Subject Line Example	Trigger Principle
IT / Security alert	"Mandatory: Password reset required by Friday"	Authority + Urgency
HR / Payroll	"Updated benefits enrollment — action needed"	Urgency + Self-interest
Shared document	"Q4 Strategy Deck — Comments requested"	Authority + Reciprocity
Delivery / Package	"Your package could not be delivered"	Urgency + Curiosity
Vendor / Partner	"Updated invoice attached — NET 30 terms"	Authority + Commitment
Conference follow-up	"Great meeting you at RSA — slides attached"	Liking + Reciprocity
Onboarding	"Welcome! Complete your setup in the employee portal"	Authority + Commitment

Notes: HR and IT pretexts consistently achieve the highest click-through rates in red team assessments.

Business Email Compromise Example

BEC is the most financially damaging form of phishing. FBI reported \$2.9 billion in BEC losses in 2023 alone.

From: David Park, CFO <d.park@nexus-c0rp.com> [typosquatted domain]

To: Maria Santos, Accounts Payable <m.santos@nexus-corp.com>

Subject: Confidential — Urgent wire transfer needed

Maria,

I need you to process an urgent wire transfer for the acquisition we discussed in the board meeting last week. This is time-sensitive and confidential — please do not discuss with other team members until the deal closes.

Amount: \$287,500.00

Beneficiary: Meridian Consulting Group

Account: [redacted]

Routing: [redacted]

I'm in back-to-back meetings until 5 PM — please process today and confirm via email. I'll approve the PO retroactively.

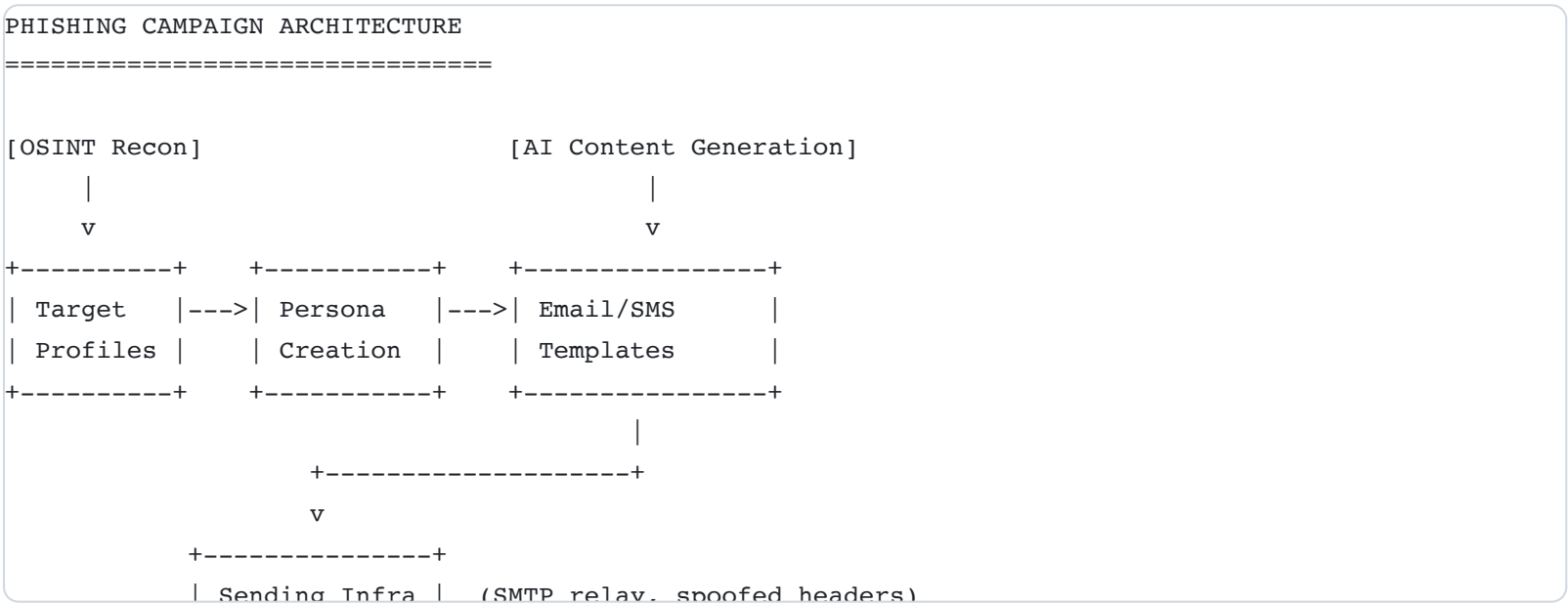
Thanks for handling this quickly.

David

Red flags: Typosquatted domain, urgency, secrecy request, process bypass, authority pressure.

Full Phishing Campaign Architecture

A professional red team phishing campaign involves multiple coordinated components.



GoPhish: Campaign Setup

GoPhish is the industry-standard open-source phishing framework for red team assessments.

```
# Install GoPhish (authorized testing only)
wget https://github.com/gophish/gophish/releases/latest/download/gophish-v0.12.1-linux-64bit.zip
unzip gophish-v0.12.1-linux-64bit.zip
cd gophish
./gophish
```

Campaign Components in GoPhish

- **Sending Profile** — SMTP configuration for sending phishing emails
- **Email Template** — HTML email with tracking pixel and phishing link
- **Landing Page** — Credential harvesting page (import from URL or custom HTML)
- **User Group** — Target list with names and email addresses
- **Campaign** — Ties all components together with scheduling

```
// GoPhish API — Create campaign programmatically
{
  "name": "Q1 Security Assessment - Nexus Corp",
```

```
"template": { "name": "IT Password Reset" },
"page": { "name": "SSO Login Clone" },
"smtp": { "name": "Assessment SMTP Relay" },
"url": "https://nexus-corp-sso.example.com",
"groups": [{ "name": "Engineering Team" }],
"launch_date": "2025-03-15T09:00:00+00:00"
}
```

Email Header Manipulation

Understanding email headers is essential for both crafting and detecting phishing.

Key Headers Attackers Manipulate

Header	Purpose	Attack Technique
From:	Displayed sender	Display name spoofing
Reply-To:	Where replies go	Route replies to attacker inbox
Return-Path:	Bounce address	Mismatched envelope sender
X-Mailer:	Sending software	Spoof to match legitimate client
Message-ID:	Unique message ID	Forge to match target domain

```
# Suspicious email headers – spot the red flags
From: "David Park, CFO" <d.park@nexus-c0rp.com>      <-- typosquatted domain
Reply-To: d.park.cfo@gmail.com                        <-- external reply address
Return-Path: bounces@mail-relay-372.com               <-- unrelated bounce domain
Received: from mail-relay-372.com (192.168.1.50)      <-- suspicious relay
```

X-Mailer: Microsoft Outlook 16.0

<-- spoofed mailer

Authentication-Results: spf=fail; dkim=none; dmarc=fail <-- FAILS all checks

Phishing Detection: The Defender's View

Understanding how defenders detect phishing makes you a better red teamer — and vice versa.

Technical Controls

Email Authentication

- **SPF** — verifies sender IP is authorized
- **DKIM** — cryptographic signature on email
- **DMARC** — policy for SPF/DKIM failures
- **BIMI** — brand logo verification

Content Analysis

- **ML classifiers** — trained on phishing corpora
- **URL reputation** — domain age, certificate analysis
- **Sandbox detonation** — execute attachments safely
- **AI detection** — identify LLM-generated text

Human Controls

- **Security awareness training** — regular phishing simulations

- **Phish reporting button** — one-click reporting in email clients
 - **Verification procedures** — out-of-band confirmation for financial requests
 - **External email banner** — visual warning for messages from outside the org
-

Bypassing Defenses with AI

As a red teamer, you need to understand how AI helps evade each layer of defense.

Defense Layer	AI Bypass Technique
Spam filters	AI generates unique text per recipient — no bulk signatures to match
URL reputation	Use legitimate services (Azure Blob, Google Docs) as redirect hops
Content ML classifiers	AI-generated text does not match known phishing templates
Grammar-based detection	LLMs produce flawless, native-quality prose
External email banners	Compromise internal accounts or use reply-chain hijacking
User awareness training	Hyper-personalization makes emails indistinguishable from real ones

The Arms Race

AI is simultaneously the best weapon for attackers and the best tool for defenders. The side that adapts faster wins.

Indicators of Phishing

Train yourself to spot these indicators — even in AI-generated messages.

Technical Indicators

- **Domain mismatch** — hover over links, check actual domain
- **SPF/DKIM/DMARC failures** — check email headers
- **Recently registered domain** — check WHOIS creation date
- **Mismatched Reply-To** — replies go somewhere unexpected
- **Suspicious attachments** — .html, .hta, macro-enabled Office docs

Behavioral Indicators

- **Unusual request** — wire transfers, credential entry, software installs
- **Urgency pressure** — "Do this now" without reasonable explanation
- **Secrecy request** — "Don't tell anyone about this"
- **Process bypass** — "Skip the usual approval chain"

- **Emotional manipulation** — fear, excitement, flattery, or guilt
-

Ethical & Legal Framework

Social engineering testing operates in a legal and ethical minefield. Clear boundaries are non-negotiable.

Legal Requirements

- **Written authorization** — scope document signed by an authorized stakeholder
- **Defined scope** — which employees, which methods, which timeframe
- **Data handling** — how captured credentials will be stored and destroyed
- **Incident response plan** — what happens if a target reports real distress
- **Jurisdictional compliance** — laws vary by country and state (CFAA, GDPR, etc.)

Ethical Guidelines

- **No personal harm** — do not exploit personal tragedies, health, or family
- **No public shaming** — report aggregate results, not individual failures

- **Proportional testing** — match intensity to the assessment's goals
 - **Debrief targets** — education is the goal, not humiliation
 - **Secure the data** — captured credentials are high-risk data; treat them accordingly
-

Rules of Engagement: Template

Every social engineering engagement must have a formal rules of engagement document.

SOCIAL ENGINEERING RULES OF ENGAGEMENT

=====

Client: [Fictional Nexus Corp]
Assessor: [Your Red Team]
Date Range: [Start] to [End]
Authorization: Signed by [Name, Title] on [Date]

SCOPE

- Phishing: Email-based pretexts targeting Engineering dept (50 users)
- Vishing: NOT authorized
- Physical: NOT authorized
- Smishing: NOT authorized

AI-Specific Ethical Concerns

AI introduces new ethical considerations beyond traditional social engineering testing.

Emerging Risks

- **Deepfake voice/video** — cloning someone's voice or face without consent, even in testing, raises serious concerns
- **Scale of deception** — AI can target thousands simultaneously; collateral damage increases
- **Persistent personas** — AI-maintained fake identities can cause lasting harm to trust networks
- **Data poisoning** — OSINT gathered for phishing may include sensitive personal information
- **Dual-use research** — techniques taught for defense are immediately usable for offense

The Red Teamer's Oath

I test systems to make them stronger, not to exploit the people who use them. I handle every credential, every piece of personal data, and every moment of someone's trust as if it were my own.

Notes: Discuss where each student draws their personal ethical line. There are no universal answers, but there must be deliberate thinking.

Defense Recommendations

After running a phishing assessment, red teams should provide actionable defense recommendations.

Technical Defenses

- Enforce DMARC with `p=reject`
- Deploy phishing-resistant MFA (FIDO2/WebAuthn)
- Implement email link rewriting and sandbox detonation
- Enable external email banners
- Monitor for lookalike domain registrations
- Deploy AI-based email analysis tools

Human Defenses

- Regular phishing simulations with progressive difficulty
- Out-of-band verification for financial requests
- No-blame reporting culture
- Role-specific training (finance, HR, executives)
- Gamified security awareness programs
- Rapid response playbooks for reported phish

Measuring Campaign Effectiveness

Quantify results to demonstrate risk and justify security investment.

Key Metrics

Metric	What It Measures	Industry Average
Email open rate	How many targets opened the email	50-70%
Click-through rate	How many clicked the phishing link	10-30%
Credential submission	How many entered credentials	5-15%
Report rate	How many reported the phish	5-15%
Time to first click	How quickly the first target fell	1-5 minutes
Time to first report	How quickly someone flagged it	15-60 minutes

Reporting Formula

Risk = (Click-through rate) x (Credential submission rate) x (Number of targets)

A 20% click rate and 10% submission rate across 500 employees = 10 compromised accounts per campaign.

Case Study: AI-Enhanced Spear Phishing

A fictional walkthrough of a complete AI-assisted social engineering engagement.

Target: Fictional Nexus Corp — Engineering Department

1. **Recon (Day 1-2):** Scraped LinkedIn for 50 engineers; identified tech stack (K8s, Go, AWS); found 3 recent hires
2. **Persona creation (Day 2):** AI generated "DevSecOps consultant" persona with full LinkedIn profile
3. **Pretext development (Day 3):** AI crafted 5 email variants — IT password reset, Kubernetes security alert, onboarding portal, shared doc, conference follow-up
4. **Infrastructure (Day 3-4):** Registered nexus-corp-sso.co; deployed GoPhish; cloned SSO login page
5. **Launch (Day 5):** Sent personalized emails to 50 targets over 4 hours (staggered to avoid bulk detection)
6. **Results (Day 5-7):** 62% open rate, 28% click-through, 14% credential submission, 8% reported

Notes: Walk through each step and ask students what they would do differently. Discuss how the 8% report rate compares to the 14% compromise rate.

CTF Challenge: Phishing Campaign Design

Educational Use Only

Social engineering and phishing techniques are demonstrated for authorized red team engagements and CTF competitions ONLY. Phishing real individuals without explicit authorization is illegal and causes real harm. All examples use fictional targets.

Scenario

You are conducting an authorized red team assessment against **Pinnacle Health Technologies** (fictional). Your goal: design a targeted phishing campaign against their engineering team using AI-assisted techniques.

Target Organization Profile

- **Company:** Pinnacle Health Technologies (fictional healthcare SaaS)

- **Size:** 200 employees, 40 engineers
 - **Tech stack:** Python, React, AWS, Docker, Slack
 - **Recent events:** Just acquired a smaller startup; migrating to new SSO provider
 - **Key personnel:** CTO — Dr. Ravi Patel; VP Eng — Lisa Chang; CISO — Marcus Webb
-

CTF Challenge: Requirements

Deliverables

1. **OSINT Report** — Document what intelligence you gathered from the fictional org profile and how it informs your campaign
2. **3 Phishing Email Templates** — Each using different pretexts and Cialdini principles; personalized for specific roles
3. **Landing Page HTML** — A credential harvesting page that matches the fictional org's branding
4. **Campaign Plan** — Timing, sending infrastructure, target selection, and escalation procedures
5. **Rules of Engagement** — Define scope, boundaries, and data handling
6. **Defense Report** — Recommend technical and human controls based on the attack you designed

Scoring Criteria

Criteria	Points
Pretext creativity and realism	25
Personalization using OSINT	20
Technical infrastructure design	20
Defense recommendations quality	20
Ethical considerations and RoE	15

CTF Challenge: Hints & Tips

Think Like an Attacker

- **Exploit the SSO migration** — new systems create confusion and urgency
- **Target new employees** — they do not know normal procedures yet
- **Use the acquisition** — "action required for system consolidation"
- **Healthcare compliance** — HIPAA training pretexts are highly effective
- **Slack as a vector** — consider multi-channel attack chains

Think Like a Defender

- What would make your own phishing emails detectable?
- How would you train users to spot your specific pretexts?
- What technical controls would block your attack chain?
- How would phishing-resistant MFA change your approach?

Remember

The best red teamers build better defenses. Your campaign should teach the organization something valuable about their security posture.

Key Takeaways

- **Social engineering targets humans, not systems** — Cialdini's principles are the exploit primitives
- **AI eliminates the quality gap** — phishing emails are now indistinguishable from legitimate communication
- **OSINT enables personalization at scale** — spear phishing is no longer expensive
- **Voice and SMS attacks are accelerating** — deepfake audio is commercially available today
- **Infrastructure matters** — domains, landing pages, and MFA bypass tools form the attack chain
- **Defense requires both technical and human controls** — neither alone is sufficient
- **Ethics are non-negotiable** — written authorization, defined scope, data protection, no personal harm
- **Red teams exist to improve defenses** — offense informs defense; the goal is always to make things better

Social engineering and phishing techniques are demonstrated for authorized red team engagements and CTF competitions ONLY. Phishing real individuals without explicit authorization is illegal and causes real harm. All examples use fictional targets.

Resources & Further Reading

Books

- **"Social Engineering: The Science of Human Hacking"** — Christopher Hadnagy
- **"The Art of Deception"** — Kevin Mitnick
- **"Influence: The Psychology of Persuasion"** — Robert Cialdini

Tools

- **GoPhish** — Open-source phishing framework (getgophish.com)
- **Evilginx2** — MFA bypass reverse proxy (github.com/kgretzky/evilginx2)
- **dnstwist** — Domain permutation engine (github.com/elceef/dnstwist)
- **theHarvester** — Email and subdomain OSINT (github.com/laramies/theHarvester)

Standards

- **PTES** — Penetration Testing Execution Standard (social engineering section)

- **NIST SP 800-115** — Technical Guide to Information Security Testing

Notes: Next lesson we move into post-exploitation. The phishing skills from this lesson become the initial access vector for everything that follows.
