# CIS 350: AI for Cybersecurity

# Week 3 Activity Worksheet: Phishing and Social Engineering

## Activity Overview

In this activity you will complete a beginner phishing lesson, then answer questions that connect phishing and social engineering to the classifier you build in Lab 3. You will learn the email features a spam classifier uses, how attackers evade the clue a detector trusts most, and why an overfit decision tree is easy to fool. Finally, you will design a practical anti-phishing checklist for a company.

### Warm-Up (5 min): Predict First

Before you study, make a quick prediction and discuss it with a partner. Do NOT look anything up yet - this is just your gut instinct.

An email company builds a 'smarter' spam filter that memorizes every phishing email it has ever seen, down to the exact wording and links. Will it catch TOMORROW's phishing better than a simpler filter that learns only general patterns? Predict yes or no in one or two sentences and say why. Jot your prediction below, then compare with your partner. You will revisit this exact idea in Task 3 (overfitting).

### Task 1: Complete the Phishing Lesson

**Resource:** Hacker101 phishing lesson (hacker101.com/lessons) or CISA Phishing Guidance (cisa.gov/phishing)
Duration: ~15 minutes

As you go, take notes on:
- How attackers make a message look trustworthy (fake senders, logos, urgency)
- Which clues give a phishing email away
- What social engineering is and why humans are the weakest link
- What defenses stop the attack before a human clicks

**Notes space:**

## Task 2: Guided Questions

Q1: In your own words, what is phishing and how is it different from social engineering?

Q2: Why are humans often called the 'weakest link' in cybersecurity? Give one example.

Q3: A spam classifier uses features like num_links, attachment, sender_mismatch, and urgent_words. Pick two and explain what each one signals about a possible phishing email.

Q4: If sender_mismatch is the strongest phishing clue, how might an attacker deliberately evade it? (Hint: think about look-alike domains.)

Q5: Name one email feature an attacker CANNOT easily spoof (for example domain age or SPF/DKIM) and explain why it resists faking.

## Task 3: Analyze Your Lab 3 Findings

In Lab 3 you build a decision-tree spam classifier, compute precision and recall, and rank feature importance. Recall the key results:
- As the tree depth grows, test accuracy plateaus and then DROPS (this is overfitting)
- Each feature gets an importance score; one feature ranks highest
- Precision and recall measure two different kinds of mistakes

**1.** Which feature ranked as the MOST important in your spam classifier? Why would an attacker deliberately target that exact feature to get their email through?

**2.** Your test accuracy rose, plateaued, and then dropped as tree depth increased. Explain what overfitting means here, and why a very deep tree can be evaded by an attacker who only copies historical spam patterns.

**3.** This spam filter protects many users at once. Discuss the precision-vs-recall trade-off: what goes wrong if the filter has high recall but low precision? What goes wrong if it has high precision but low recall? Which would you favor, and why?

## Task 4: Design an Anti-Phishing Checklist

You are a security analyst at a mid-size company. Employees keep clicking phishing links. Design a practical anti-phishing checklist or policy with 5-7 items that a non-technical employee could actually follow before trusting an email.

**Your checklist (use a numbered list):**

**Explain your choices:** For two of your items, explain which attack technique from the lesson it defends against and why it works.

## Task 5: Reflection

Q1: What surprised you most about how attackers trick people, and why?

Q2: How does thinking about phishing as a classifier problem (phishing vs. safe) change the way you understand overfitting from this week's lecture?

Q3: What is one habit you will change about how you handle suspicious emails after this activity?

## Grading Rubric (20 points)

| Criterion | Description | Points |
|---|---|---|
| Completion | Lesson completed; all 5 tasks and questions attempted | 8 |
| Understanding | Q&A answers demonstrate learning about phishing and classifiers | 7 |
| Application | Anti-phishing checklist is practical, clear, and well-explained | 5 |

**Total: 20 points**

Scoring: Completion (8) + Understanding (7) + Application (5) = 20 points

## How to Submit

1. Complete all tasks in this worksheet
2. Ensure all questions are answered thoughtfully
3. Submit this PDF on Canvas by [DUE DATE]

Questions? Post in the course forum or attend office hours!