HANDOUT 13 - WEEK 15 READING

Advanced Topics and Future Threats

CIS 350: AI for Cybersecurity - Smart Defense for the Digital Business

Colorado State University - Computer Information Systems

How to use this reading

This handout supports Week 15 - the last new-content week. It covers advanced ML techniques that help when labeled data is scarce (transfer, few-shot, and federated learning), the emerging AI-enabled threats on the horizon, and how the field is likely to evolve. The figures match the lecture; the review questions map to Quiz 12 and help you prepare for Exam 4 (Weeks 10-16).

Estimated reading time: 20-25 minutes. No prior coding background needed.

1. The Problem These Techniques Solve: Scarce Labels

A theme runs through this whole course: in security you usually have LOTS of data but very FEW labels. Labeling attacks is slow, expensive, and needs experts - and tomorrow's threat may have zero examples today. Three advanced techniques squeeze more out of few (or no) labels: transfer learning, few-shot learning, and federated learning.

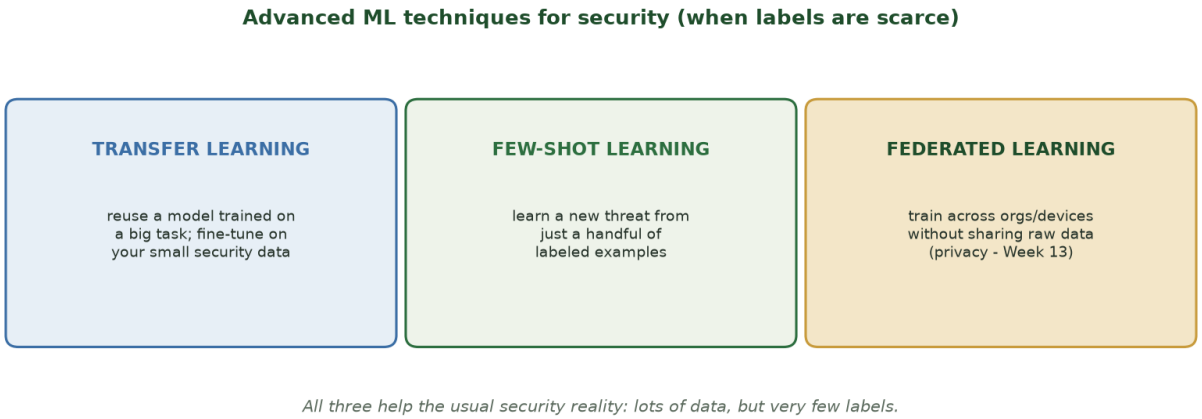

Figure 1. Three advanced techniques, all aimed at the 'lots of data, few labels' problem.

2. Transfer & Few-Shot Learning

TRANSFER LEARNING starts from a model already trained on a huge general task and FINE-TUNES it on your small, specific dataset. Because the pretrained model already learned useful general features, your limited data goes much further - like a trained analyst who ramps up on your network far faster than a novice. FEW-SHOT LEARNING pushes this to the extreme: recognize a new class or threat from just a handful of examples (ZERO-SHOT does it from a description alone). Few-shot usually builds on a strong pretrained base.

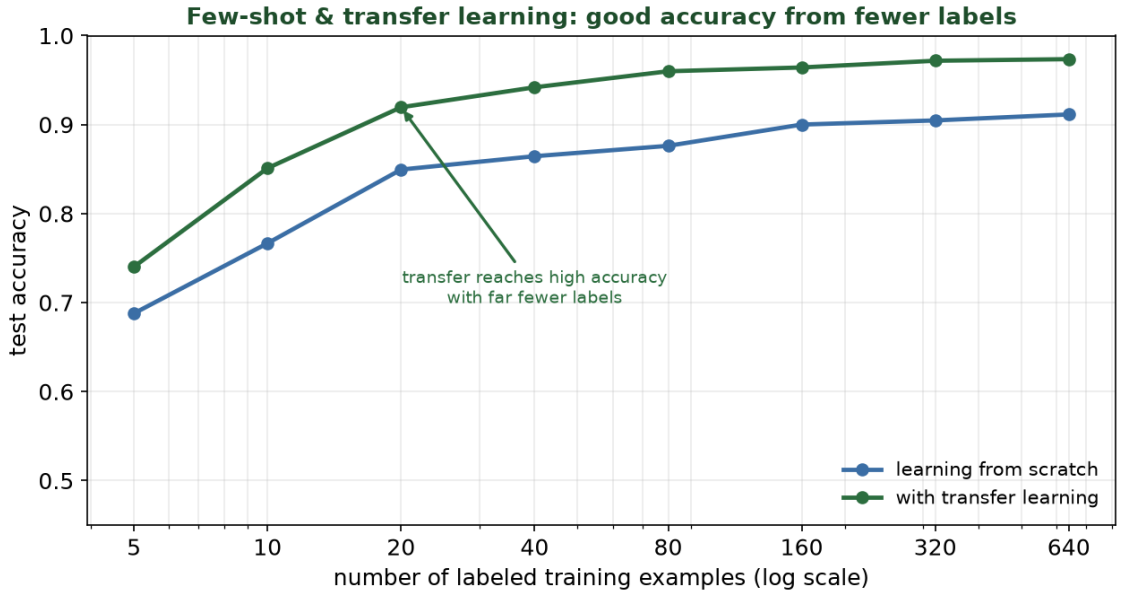

Figure 2. A real learning curve: with transfer, ~20 labels reach ~0.92 accuracy - matching 'from scratch' at ~640 labels (about 30x fewer labels for the same accuracy).

Why this matters in security

Novel attacks appear with few examples. Transfer and few-shot learning let you build a useful detector from a handful of labels instead of thousands - the difference between reacting in days vs. months.

3. Federated Learning (Recall Week 13)

FEDERATED LEARNING trains a shared model across many organizations or devices WITHOUT sharing raw data: each site trains locally and sends only model updates, which are combined centrally. In security, this lets banks or hospitals jointly detect threats none of them sees enough of alone - while keeping customer data private. It is often paired with differential privacy for stronger protection.

4. Emerging AI-Enabled Threats

AI is dual-use: every capability we use for defense can be turned to offense (recall Week 8). The emerging threats push automation and scale further. We study them ONLY to defend - never to build them.

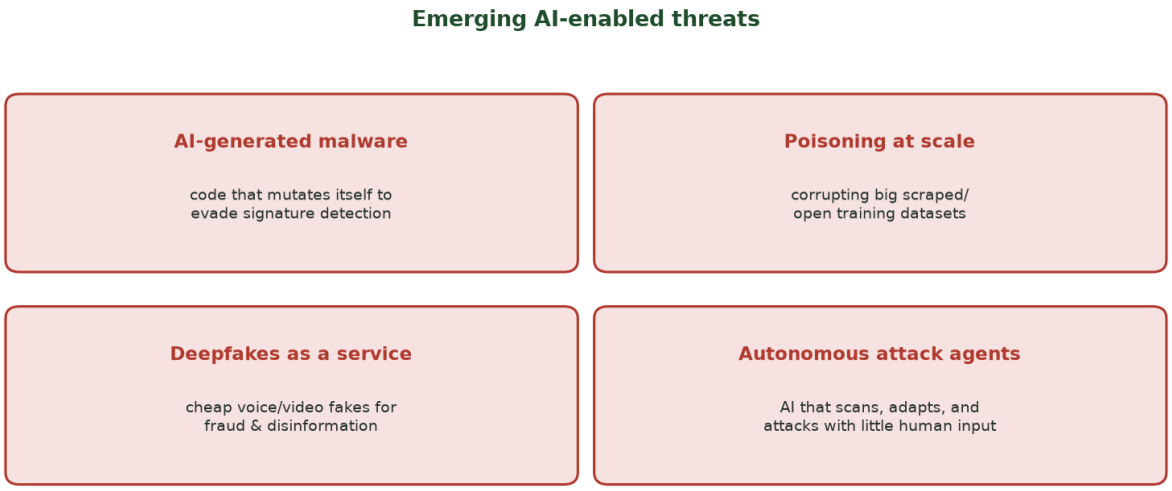

Figure 3. Four emerging threats: AI-generated malware, poisoning at scale, deepfakes-as-a-service, and autonomous attack agents.

- AI-generated / polymorphic malware: mutates to evade signature detection -> use behavior/anomaly detection, not signatures alone.
- Poisoning at scale: seeding malicious content into big scraped/open training data -> vet data, track provenance, curate sources, test for backdoors.
- Deepfakes as a service: cheap voice/video fakes for fraud -> out-of-band verification and MFA.
- Autonomous attack agents: AI that scans, adapts, and attacks with little human input -> monitoring, rate-limiting, and defense in depth.

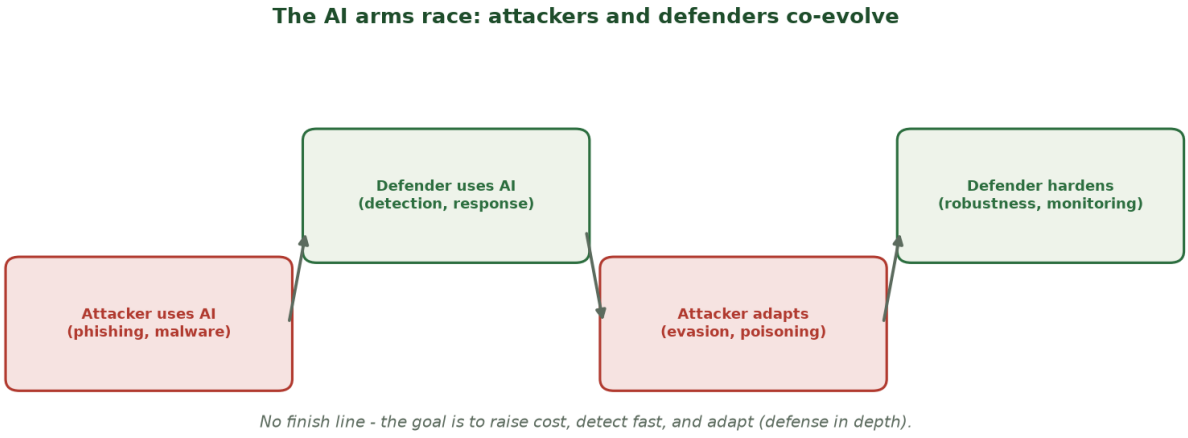

Figure 4. The arms race: attackers and defenders keep adopting AI and adapting to each other.

There is no finish line

Because attackers and defenders co-evolve, the goal is never a permanent 'win' - it is to raise the attacker's cost, detect fast, and adapt. That is defense in depth, the spine of this course.

5. The Future of AI + Cybersecurity

Expect more automation on both sides, AI 'agents' that ACT (not just advise), and tighter regulation (the EU AI Act and successors). As AI acts more, human oversight and least privilege matter MORE, not less. The reassuring part: the fundamentals you learned this term - defense in depth, trade-offs, trustworthy AI, and keeping a human in the loop - are exactly what keep you effective as the tools change.

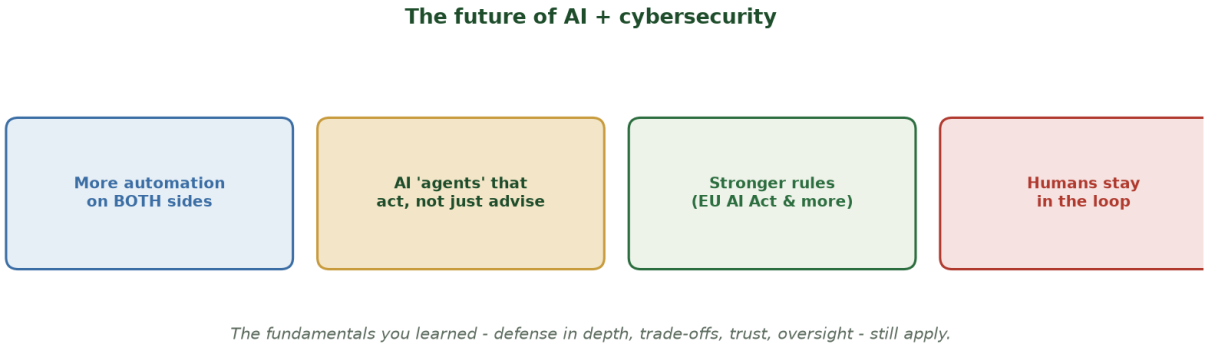

Figure 5. Trends to watch - and the fundamentals that endure through all of them.

6. Where to Go From Here

- Roles: security analyst, AI/ML security engineer, AI governance/GRC, risk & audit.
- You built a real portfolio: 12 hands-on labs, a beginner CTF, and a tabletop exercise.
- Stay current: follow NIST, CISA, OWASP, and MITRE ATLAS; read annual threat reports.
- Keep practicing: CTFs, Kaggle, open datasets, and reproducing papers.
- Fundamentals + curiosity + ethics travel with you as the field evolves.

Key Terms

- Transfer learning - reuse a pretrained model; fine-tune on your small dataset.
- Few-shot / zero-shot learning - learn from a handful of examples / from a description alone.
- Federated learning - shared model, private data (only updates are shared).
- Label scarcity - the security reality of lots of data but few labels.
- Polymorphic malware - malware that mutates to evade signatures.
- Poisoning at scale - corrupting large scraped/open training datasets.
- Deepfake-as-a-service; autonomous attack agent - emerging AI-enabled threats.
- AI arms race - attackers and defenders continually adapting to each other.

Review Questions (self-check for Quiz 12 and Exam 4)

- What problem do transfer, few-shot, and federated learning all address?
- Why does transfer learning need fewer labels than training from scratch?
- Name two emerging AI-enabled threats and one defense for each.
- Explain the AI arms race and the defensive goal (since there is no permanent win).
- Name two course fundamentals that will still matter as AI and threats evolve, and why.

Remember: AI tools are not permitted on quizzes or exams. Tools change fast; the fundamentals you learned endure.

References

- S. J. Pan & Q. Yang, 'A Survey on Transfer Learning,' IEEE TKDE, 2010.
- Y. Wang et al., 'Generalizing from a Few Examples: A Survey on Few-Shot Learning,' 2020.
- H. B. McMahan et al., 'Communication-Efficient Learning of Deep Networks' (federated learning), 2017.
- ENISA Threat Landscape reports; MITRE ATLAS; NIST AI RMF 1.0 (2023).
- CISA/NCSC, Guidelines for Secure AI System Development, 2023.
- Chio & Freeman, 'Machine Learning and Security' (O'Reilly).