EXAM 3 PRACTICE - WEEKS 10-13

Practice Exam

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

Colorado State University - Computer Information Systems

About this document

This is a PRACTICE exam for Exam 3. The questions differ from the real exam but cover the same Weeks 10-13 material in the same format and length (100 points; 90 minutes): multiple choice, true/false, short answer, and applied scenarios. Try it closed-book and timed, then check the separate answer key. Exam 3 is taken online in Canvas with LockDown Browser.

Total: 100 points, 90 minutes. Part A multiple choice (20 x 2 = 40), Part B true/false (10 x 1 = 10), Part C short answer (5 x 4 = 20), Part D applied scenarios (3 x 10 = 30). No heavy math, no coding.

Part A - Multiple Choice (20 x 2 = 40 pts)

1. Which best describes an adversarial example?

A. A large, obvious edit anyone would notice
B. A labeled training dataset
C. A tiny, crafted change to an input that flips the prediction
D. A network port scan

2. A backdoor attack is best described as:

A. Poisoning that plants a hidden trigger for secret misbehavior
B. A type of firewall
C. A faster encryption method
D. An evasion attack at test time

3. A membership-inference attack tries to learn:

A. The model's speed
B. The number of layers
C. The GPU model
D. Whether a specific record was in the training set

4. The best overall mindset for defending AI systems is:

A. Rely on one perfect tool
B. Defense in depth
C. Ignore attacks
D. Never deploy anything

5. Model extraction (theft) works by:

A. Heavily querying a model to clone its behavior
B. Deleting the model
C. Adding labels to data
D. Speeding the model up

6. The root cause of prompt injection is that an LLM:

A. Runs offline
B. Has too few parameters
C. Understands meaning perfectly
D. Cannot reliably separate trusted instructions from untrusted data

7. A jailbreak attack tries to:

A. Speed up the model
B. Get the model past its safety rules
C. Back up the prompt
D. Encrypt the output

8. Prompt leakage happens when the model:

A. Runs slowly
B. Adds random noise
C. Reveals its hidden system prompt
D. Deletes data

9. The most important architectural control for an LLM that can use tools is:

A. A bigger model
B. More training data
C. A faster GPU
D. Least privilege

10. An output guardrail:

A. Checks the model's response before it is shown or acted on
B. Speeds up responses
C. Is just part of the prompt
D. Stores passwords

11. We disaggregate metrics by group because overall accuracy:

A. Is always wrong
B. Is illegal
C. Can hide per-group harm
D. Needs no data

12. Equal opportunity (a fairness definition) means equal:

A. File sizes
B. True-positive rate across groups
C. Numbers of features
D. Training time

13. The EU AI Act regulates AI systems by:

A. Risk level (higher risk -> stricter duties)
B. Color
C. File size
D. Alphabetical order

14. A bias 'feedback loop' means:

A. Training gets faster
B. You add more GPUs
C. Biased outputs become tomorrow's training data
D. You use fewer features

15. Common fairness definitions:

A. Are all identical
B. Can conflict, so fairness needs a deliberate choice
C. Never matter in practice
D. Are illegal

16. Making the differential-privacy budget epsilon SMALLER gives:

A. Less privacy
B. No effect
C. A faster GPU
D. More privacy but usually lower accuracy

17. In federated learning, what is sent to the central server?

A. All the raw data
B. Nothing, ever
C. Only the model updates
D. The user's password

18. Homomorphic encryption lets you:

A. Compute on encrypted data without decrypting it
B. Skip training
C. Delete data
D. Speed up the GPU

19. Anonymizing by deleting names is weak because:

A. It is illegal
B. It is slow
C. It encrypts the data
D. Combining other fields can re-identify people

20. Collecting more security data:

A. Has no privacy cost
B. Helps detection but raises privacy risk
C. Is always illegal
D. Slows the internet

Part B - True / False (10 x 1 = 10 pts)

21. Data poisoning corrupts the training data. (T / F)
22. An adversarial example must be a large, obvious change to the input. (T / F)
23. Prompt injection is the #1 risk on the OWASP Top 10 for LLM Applications. (T / F)
24. A model can be accurate overall yet unfair to a group. (T / F)
25. A strong system prompt alone stops all prompt injection. (T / F)
26. Removing the sensitive attribute always removes bias. (T / F)
27. In differential privacy, a smaller epsilon means more privacy. (T / F)
28. Federated learning centralizes all raw data on one server. (T / F)
29. There is one perfect defense that solves all AI security problems. (T / F)
30. Strong encryption can help privacy but make defender inspection harder. (T / F)

Part C - Short Answer (5 x 4 = 20 pts)

31. What is data poisoning, and how does it differ (in timing) from an adversarial/evasion attack?

32. What is prompt injection, and name two defenses for an LLM application.

33. Why can a model be accurate overall but unfair, and name one fairness definition.

34. What does epsilon control in differential privacy, and what is the trade-off?

35. Compare federated learning and encrypted computation in one sentence each.

Part D - Applied Scenarios (3 x 10 = 30 pts)

36. A hospital readmission model (10 pts)

A hospital uses ML to predict which patients will be readmitted, then targets follow-up care.

(a) An audit finds the model under-serves one group. Which per-group metric would reveal this? (3 pts)
(b) Patients worry the model could leak their records. Name a privacy attack and one defense. (3 pts)
(c) Several hospitals want a shared model without pooling raw patient data. Which technique? (2 pts)
(d) Name one governance step before deployment. (2 pts)

37. A company AI code assistant (10 pts)

A firm deploys an LLM assistant that reads internal docs and can run limited commands.

(a) Name the top risk and one variant (direct or indirect injection). (2 pts)
(b) Why is a good system prompt not sufficient? (2 pts)
(c) Give one guardrail and one least-privilege decision. (4 pts)
(d) Name one monitoring or human-oversight control. (2 pts)

38. A malware detector under attack (10 pts)

A security vendor ships an ML malware detector that attackers actively try to defeat.

(a) Attackers tweak malware to slip past it. Name the attack and one defense. (3 pts)
(b) They also try to corrupt the training feed. Name that attack and one defense. (3 pts)
(c) The team wants to collect lots of customer telemetry. Name the privacy concern and a middle ground. (2 pts)
(d) Why is there 'no perfect defense,' and what should the team do? (2 pts)

End of practice exam. Check your answers with the separate answer key.