PRACTICE QUESTIONS

Linux Foundation / CNCF

SAIA

Secure AI Associate

PUBLISHED BY aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
ISSUED	November 8, 2026
EXAM	Linux Foundation / CNCF SAIA, Secure AI Associate

Shared by **aiexamcerts** on aiexamcerts

10 practice questions for the Linux Foundation / CNCF SAIA exam, Secure AI Associate, each followed by its answer and a short explanation. More free sets and the full practice exam: **aiexamcerts.com/linux-foundation-cncf/saia**. Spotted a mistake? Email support@aiexamcerts.com with the exam code SAIA and the question number.

---

1. When starting a threat model for a new ML-enabled service, what is the best first step?

A Pick a noise-injection defense and apply it to the model before any analysis

B Run a fuzzer against the production inference endpoint and review the crashes

C Increase the training set size so the model accuracy improves on held-out data

D Diagram data flows and trust boundaries across training, artifacts, and serving

Answer: D

Threat modeling begins with understanding assets, data flows, and trust boundaries so threats can be enumerated against them. Selecting a defense first, such as noise injection, skips the analysis that tells you which defense is needed.

---

2. A training set contains customer names and phone numbers that the model does not need. What is the best action?

A Encrypt only the model output and leave the raw training records as they are

B Minimize the data by removing or de-identifying it before training

C Share the raw set with all engineers so they can decide what to remove later

D Keep everything in the set because more data always improves the final model

Answer: B

Data minimization reduces privacy exposure at the source. Keeping unneeded personal data increases both leakage risk and compliance burden.

---

3. Giving an agent more functionality, permissions, or autonomy than it needs is called excessive _____.

Answer: agency

Excessive agency enlarges the damage possible if the model is manipulated or errs.

---

4. The person accountable for a model's risk decisions and incident escalation is its _____.

Answer: owner (also accepted: model owner)

A named owner makes accountability explicit across the lifecycle.

---

5. A model performs much worse on a demographic group that is underrepresented in training data. Which practice addresses this?

A Remove the evaluation set so subgroup results can no longer be computed

B Report only the overall accuracy figure to stakeholders in the release notes

C Increase the batch size used in training to smooth out the group differences

D Check representativeness, rebalance the data, and evaluate by subgroup

Answer: D

Representativeness checks and subgroup evaluation surface and reduce disparity. Overall accuracy can hide poor performance on a minority group.

---

6. An assistant summarizes web pages. A page hides text telling it to email the user's contacts to an attacker. Which mitigation is most effective?

A Increase the model's sampling temperature so the injected text is less likely to be followed

B Cache every page the assistant has ever read so that the content can be replayed later

C Treat retrieved text as untrusted, limit tool privileges, and confirm sensitive actions

D Add a system prompt line saying the assistant must never follow any hidden text it reads

Answer: C

This is indirect prompt injection, and layered controls on privileges and confirmations contain it. A prompt instruction alone is bypassable, which makes it the tempting but insufficient answer.

---

7. Which measure best reduces overreliance on an AI assistant's output for high-impact decisions?

A Telling users in onboarding material that the model is always correct

B Hiding all sources and confidence cues so that answers look cleaner to users

C Human review of high-impact outputs, with citations shown to users

D Disabling the review step on high-impact decisions in order to save time

Answer: C

Human oversight and transparency let people check results before acting. Hiding sources removes the information users need to verify.

---

8. An LLM proposes a function call with arguments. What should the application do before executing it?

A Execute it immediately because the model's output is considered trusted by default

B Run it with the service account of the highest privilege to avoid permission errors

C Validate arguments against a schema and authorize under the user's permissions

D Skip validation whenever the call pattern was suggested by the system prompt

Answer: C

Model-proposed calls are untrusted input and need validation and authorization like any other request. Using elevated privileges would allow privilege escalation.

---

9. Which information should experiment tracking record so a model can be reproduced and audited?

A The colors and chart styles that were chosen for the team dashboards

B Code commit, dataset version, hyperparameters, and evaluation metrics

C Only the final accuracy number, written once into the release announcement

D The wallpaper and theme the developer used on their desktop environment

Answer: B

Linking code, data, parameters, and results makes a model traceable. A lone accuracy figure cannot be reproduced or audited.

---

10. What does a model card give reviewers and users of a released model?

A Setting the inference batch size and the maximum number of concurrent requests

B Storing the model's encryption keys so that the operations team can find them

C Documenting intended use, limitations, evaluation data, performance, and known risks

D Replacing the training pipeline with a single manually written script

Answer: C

Model cards support transparency and auditability by recording how a model should and should not be used. Storing keys in documentation would itself be a security flaw.