

PRACTICE QUESTIONS

CertiProf

AIAM

AI Agent Manager Professional Certification

SHARED BY [u/Fuzzy-Comet-7222](#) on aiexamcerts

EDITION Free Set 2

QUESTIONS 10

SHARED October 15, 2026

EXAM CertiProf AIAM, AI Agent Manager Professional Certification

Shared by [u/Fuzzy-Comet-7222](#) on aiexamcerts

10 practice questions for the CertiProf AIAM exam, AI Agent Manager Professional Certification, each followed by its answer and a short explanation. More free sets and the full practice exam:

www.aiexamcerts.com/certiprof/aiam. Spotted a mistake? Email support@aiexamcerts.com with the exam code AIAM and the question number.

1. Why should an agent's tool call be designed to be idempotent?
 - A The tool always returns a cached answer after its first call
 - B The tool can only be invoked by a human operator directly
 - C Repeated calls produce one effect and no duplicate side effects
 - D The tool responds faster each time it is called repeatedly in a row

Answer: C

Idempotent operations are safe to retry after a timeout because a second attempt does not create a duplicate charge, record or message. A cached answer is a tempting reading, but caching is not what idempotency guarantees. Human-only invocation and speed gains do not address retry safety.

2. An AI agent is by definition a model that has been fine-tuned on a company's private data.
 - A True
 - B False

Answer: B (False)

This is false. An agent is defined by its loop of planning, acting and observing with tools, and fine-tuning on private data is an optional choice rather than a requirement of agency.

3. What is an LLM-as-judge evaluator best used for in agent testing?
 - A Replacing all security reviews of the agent's tool permissions
 - B Setting the agent's rate limits in production traffic
 - C Storing the agent's long-term memory in a vector index
 - D Scoring open-ended outputs against a written rubric at scale

Answer: D

An LLM judge can apply a rubric to many open-ended outputs quickly, which makes large evaluation runs practical when humans spot-check the judge. It is not a substitute for security review, which is a tempting misuse. Rate limits and memory storage are separate runtime concerns.

4. Memory that holds only the information inside the current context window is called short-term or _____ memory.

Answer: working

Working memory is the material currently in the context window, and it ends with the session unless it is saved. Long-term memory persists across sessions, which is the tempting confusion.

5. Which two statements about multi-agent systems are correct? (Choose two.)

- A More agents always improve the accuracy of the final result
- B Shared state needs clear ownership to avoid conflicting writes
- C Agents can amplify each other's errors without validation between them
- D Each agent must use a different model vendor from the others
- E Coordination overhead never affects the latency of a workflow

Answer: B, C

Shared state needs an owner so that two agents do not overwrite each other, and without validation between agents an early error can spread. More agents do not automatically improve accuracy, which is the tempting belief. Vendor diversity is not required, and coordination adds latency rather than none.

6. Why should an agent's tool descriptions be precise and narrow?

- A To make the tool list shorter on the vendor invoice
- B To reduce the number of tools the vendor can bill for
- C To help the model choose the right tool and pass valid arguments
- D To avoid needing any input schema at all for the tool

Answer: C

Clear, narrow descriptions help the model select the correct tool and fill its parameters accurately, which reduces wrong calls. Shorter lists and billing counts are not the reason tool descriptions matter. Removing input schemas works against reliable argument passing.

7. A customer-support agent can issue refunds. Which design is the safest first step?

- A Give it unrestricted refund authority so resolution is faster
- B Let it refund any amount because the model reasons carefully about it
- C Hide the refund tool from the audit log to keep reports simple
- D Cap refunds at a limit and require human approval above it

Answer: D

A cap with human approval above it limits financial exposure while still letting the agent resolve routine cases. Unrestricted authority is the tempting shortcut, but it removes any check on costly mistakes. Hiding tool use from logs weakens accountability.

8. Adding more agents to a workflow always reduces its total cost and latency.

- A True
- B False

Answer: B (False)

This is false. Additional agents add model calls and coordination steps, so cost and latency often rise, which is why each agent should earn its place.

9. Which storage type is most commonly used for semantic retrieval of past notes by meaning?

- A A flat log file with no index
- B A tokenizer vocabulary file
- C A configuration YAML file for the orchestrator
- D A vector index over embeddings of the notes

Answer: D

A vector index stores embeddings so notes can be ranked by semantic similarity to a query. A flat log file can hold the same text but offers no meaning-based search without an extra index. Tokenizer vocabularies and configuration files do not store notes for retrieval.

10. Which control is the most direct defense against an agent being manipulated by instructions hidden in a web page it reads?

- A Increasing the model's temperature for more cautious answers
- B Removing the tool list from the prompt entirely
- C Treating page content as untrusted data and limiting actions
- D Asking the agent in its system prompt to ignore suspicious text

Answer: C

Treating retrieved content as data and restricting which actions can follow from it limits what an injected instruction can achieve. A system prompt request to ignore suspicious text is the tempting fix, but models can still be steered by it. Temperature and tool lists do not address the injected instruction path.