CompTIA SecAI+ CY0-001 Practice Questions

Shared by Fowler on 13-08-2026

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

Which of the following is the most impactful security risk associated with the use of a generative AI chatbot?

Options:

A- Overly permissive access
B- Data leakage
C- Weak encryption
D- Model validation

Answer:

B

Explanation:

Basic Concept: Generative AI chatbots interact with users in natural language and may access organizational knowledge bases, databases, or prior conversations. The conversational nature of these systems creates unique risks around sensitive information disclosure. CompTIA SecAI+ Study Guide ranks data leakage as the primary security concern for generative AI chatbots.

Why B is Correct: Data leakage occurs when a generative AI chatbot inadvertently reveals sensitive information including PII, confidential business data, intellectual property, training data, or system configurations in its responses. This can happen through prompt injection attacks, insufficient output filtering, or the model memorizing and reproducing sensitive training data. The impact is immediate, potentially irreversible, and can result in regulatory violations, competitive disadvantage, and reputational damage.

Why A is Wrong: Overly permissive access is a contributing factor that can exacerbate data leakage but is an access control design issue rather than the most directly impactful runtime risk of operating a generative AI chatbot.

Why C is Wrong: Weak encryption is a data protection concern for data in transit or at rest. While important, it is a configuration issue separate from the generative AI chatbot's core operational risks and is not specific to chatbot technology.

Why D is Wrong: Model validation ensures a model performs as expected before deployment. While important for quality assurance, it is a development lifecycle activity rather than an ongoing operational security risk associated with running a chatbot.

Question 2

Question Type: MultipleChoice

A security administrator wants to prevent prompt injection attacks and ensure responses have sanitized output.

Which of the following provides a main compensating control for these requirements?

Options:

A- Least privilege
B- Encryption
C- A large language model (LLM) firewall
D- Rate limiting

Answer:

C

Explanation:

Basic Concept: Preventing prompt injection and ensuring output sanitization requires a control that can inspect both the semantic content of incoming prompts and the safety of outgoing responses. This requires an intelligent, context-aware filtering layer specifically designed for LLM traffic. CompTIA SecAI+ Study Guide identifies LLM firewalls as a primary control for prompt security and output safety.

Why C is Correct: An LLM firewall is specifically designed to inspect, filter, and sanitize both incoming prompts and outgoing AI responses. It can detect and block prompt injection attempts using pattern matching, semantic analysis, and behavioral heuristics, while also sanitizing output to remove sensitive data, harmful content, or policy violations before responses reach users. This dual capability makes it the primary control addressing both requirements simultaneously.

Why A is Wrong: Least privilege restricts what resources and actions users and systems can access. It reduces the potential impact of successful attacks but does not inspect prompt content for injection attempts or sanitize model outputs.

Why B is Wrong: Encryption protects data confidentiality in transit and at rest. It does not analyze prompt content for malicious patterns or filter AI-generated responses for unsafe content. Encrypted traffic can still carry prompt injection attacks.

Why D is Wrong: Rate limiting controls request frequency. While it can slow down automated injection attack campaigns, it does not inspect the content of individual prompts to detect injections, nor does it sanitize output responses. Malicious prompts can still succeed within rate limits.

Question 3

Question Type: MultipleChoice

A cybersecurity administrator needs a security mechanism that can validate input.

Which of the following controls should the administrator use?

Options:

A- Prompt firewall
B- Rate limits
C- Token limits
D- Input quantity

Answer:

A

Explanation:

Basic Concept: Input validation is a fundamental security principle that checks incoming data against expected criteria before processing it. For AI systems, this requires a mechanism capable of inspecting the semantic content and structure of inputs --- not just their volume or format. CompTIA SecAI+ Study Guide identifies prompt firewalls as the primary input validation control for AI systems.

Why A is Correct: A prompt firewall validates incoming inputs by inspecting their content against security policies, detecting malicious patterns such as injection strings or jailbreaking attempts, enforcing structural rules, and blocking non-compliant inputs before they reach the AI model. Unlike network firewalls that operate on packet headers, a prompt firewall understands the semantic content of AI prompts, making it the appropriate input validation mechanism for AI systems.

Why B is Wrong: Rate limits control how frequently inputs are submitted, not what those inputs contain. A malicious prompt submitted within rate limits will not be detected or blocked --- rate limiting does not validate the content or intent of individual inputs.

Why C is Wrong: Token limits cap the maximum length of inputs and outputs in terms of tokens. While this can prevent excessively long inputs from being processed, it does not inspect input content for malicious patterns or validate that inputs conform to policy requirements.

Why D is Wrong: Input quantity is a generic term that might refer to limiting the number or size of inputs. Like token limits and rate limits, quantity controls do not validate the content of

inputs for security compliance or detect malicious prompt patterns.

Question 4

Question Type: MultipleChoice

Which of the following describe the practice of providing examples in a prompt? (Select two.)

Options:

A- User prompt
B- System prompt
C- Prompt template
D- Quantization
E- One-shot
F- Multi-shot

Answer:

E, F

Explanation:

Basic Concept: Prompting techniques for LLMs include various approaches to guide model behavior. Providing examples within prompts is a powerful technique that leverages the model's in-context learning capability to guide response format and quality. CompTIA SecAI+ Study Guide covers prompting techniques under basic AI concepts.

Why E is Correct: One-shot prompting involves providing exactly one example within a prompt to demonstrate to the model the desired input-output format or response style. This single example guides the model's understanding of the task without requiring extensive fine-tuning. It is a well-established prompting technique that uses examples to inform model behavior.

Why F is Correct: Multi-shot prompting (also called few-shot prompting) involves providing multiple examples within a prompt to further clarify the desired output pattern. Multiple examples help the model identify consistent patterns and produce more accurate, consistent responses. Both one-shot and multi-shot are specifically defined by their use of examples in prompts.

Why A is Wrong: A user prompt is the input message submitted by a user to the AI system. It is the general term for any user input, not a specific technique that describes the practice of providing examples.

Why B is Wrong: A system prompt sets the model's behavior, persona, and constraints at the session level. While a system prompt could contain examples, the term specifically refers to the system-level instruction context, not the technique of example provision.

Why C is Wrong: A prompt template is a reusable structured format with placeholders for variable inputs. It standardizes prompt structure but is not defined by the practice of including examples.

Why D is Wrong: Quantization is a model compression technique that reduces model size by representing weights with lower precision numbers. It is a model optimization technique completely unrelated to prompting practices.

Question 5

Question Type: MultipleChoice

An AI security team must assess the probability of an attack on its new system and the impact associated with such an attack.

Which option best threat-modeling resources best addresses the threat landscape for machine learning (ML)?

Options:

A- Common Vulnerabilities and Exposures (CVE) AI working group
B- MITRE Adversarial Threat Landscape for AI Systems (ATLAS)
C- Massachusetts Institute of Technology (MIT) risk repository
D- Open Worldwide Application Security Project (OWASP)

Answer:

B

Explanation:

Basic Concept: Assessing attack probability and impact for ML systems requires a resource specifically built to catalog real-world adversarial attacks against AI and ML systems, including documented techniques with associated impact information. CompTIA SecAI+ Exam Objectives identify MITRE ATLAS as the authoritative ML threat landscape resource.

Why B is Correct: MITRE ATLAS is specifically designed as a comprehensive knowledge base of adversarial tactics, techniques, and case studies targeting AI and ML systems. It catalogs real-world attacks with associated probability factors derived from actual incidents and provides

impact assessments for various attack types including data poisoning, model evasion, model extraction, and inference attacks. This directly enables the probability and impact assessment the team requires.

Why A is Wrong: The CVE AI working group focuses on identifying and cataloging specific vulnerability instances in AI software components. While useful for vulnerability management, it does not provide the comprehensive threat landscape coverage with probability and impact assessments for ML-specific attack tactics that ATLAS provides.

Why C is Wrong: The MIT risk repository is an academic resource cataloging general AI-related risks. It is research-oriented and does not provide the practitioner-focused, operational attack taxonomy and case study library that MITRE ATLAS offers for ML threat modeling.

Why D is Wrong: OWASP provides application security guidance including the OWASP LLM Top 10. While valuable for LLM-specific risks, OWASP does not provide the comprehensive ML threat landscape coverage or the probability and impact data that MITRE ATLAS offers for assessing the full spectrum of ML attack scenarios.

Question 6

Question Type: MultipleChoice

An organization wants to reduce vulnerabilities after deployment. The organization decides to incorporate an AI-assisted early detection and vulnerability identification process in its development workflow.

Which of the following AI-assisted functions is the best option?

Options:

A- Code linting
B- Incident management
C- Automated deployment/rollback
D- System auditing

Answer:

A

Explanation:

Basic Concept: Reducing post-deployment vulnerabilities requires catching security issues as early as possible in the development workflow. AI-assisted tools that analyze code during

development provide the earliest possible intervention point. CompTIA SecAI+ Study Guide covers AI integration in secure development under AI-assisted security.

Why A is Correct: AI-assisted code linting analyzes source code in real time during development to identify security vulnerabilities, insecure coding patterns, policy violations, and quality issues before code is compiled or committed. By catching vulnerabilities at the coding stage --- the earliest possible point in the development workflow --- AI code linting prevents vulnerable code from progressing to testing, staging, or production, directly reducing post-deployment vulnerabilities at their source.

Why B is Wrong: Incident management handles security events and incidents after they have occurred in production. It is a reactive capability focused on response and recovery rather than early-stage vulnerability identification in the development workflow.

Why C is Wrong: Automated deployment/rollback automates the process of pushing code to production and reverting to previous versions when issues are detected post-deployment. It is a deployment safety mechanism rather than an early detection tool during the development phase.

Why D is Wrong: System auditing reviews and records system activities and configurations for compliance verification. It is primarily a detective and compliance control for systems that are already deployed, not an early development-phase vulnerability identification tool.

Question 7

Question Type: MultipleChoice

SIMULATION

Instructions: Click the (+) to assign each threat category into its appropriate framework.

An architect is modeling an agentic system to meet security standards.

Options:

A- See Explanation below for complete solution for this PBQ

Answer:

A

Explanation:

MAESTRO: Overreliance, Insecure plug-in design
OWASP Top 10: Broken access control, Identification and authentication failures
OWASP Top 10 LLM: Prompt injection, Model denial of service
STRIDE: Elevation of privilege, Repudiation, Supply chain vulnerabilities, Insecure design

MAESTRO addresses AI-specific misuse such as excessive trust in outputs (overreliance) and unsafe plug-in design.

OWASP Top 10 maps to classic web threats: broken access control and authentication failures are key risks.

OWASP Top 10 LLM focuses on LLM-related threats like prompt injection and denial of service targeting the model.

STRIDE categories cover privilege escalation, repudiation (denying actions), supply chain risks, and insecure design flaws.

Basic Concept: This is a Performance-Based Question (PBQ) --- a simulation item requiring interactive drag-and-drop assignment of threat categories to appropriate frameworks in the actual exam. It tests knowledge of how different AI threat frameworks categorize and address specific threat types for agentic systems.

Key Concept --- Framework-to-Threat Mapping: MITRE ATLAS covers ML-specific adversarial tactics such as model evasion, data poisoning, model extraction, and prompt injection for

agentic systems. OWASP LLM Top 10 addresses application-level LLM vulnerabilities such as insecure output handling, excessive agency, and supply chain risks. NIST AI RMF addresses governance-level risks across the AI lifecycle. STRIDE addresses architectural threats including spoofing, tampering, repudiation, information disclosure, DoS, and elevation of privilege.

Why This Matters: Agentic AI systems have a unique threat landscape combining traditional software vulnerabilities with AI-specific attacks. Correctly mapping threat categories to frameworks is essential for comprehensive threat modeling of systems that autonomously execute multi-step tasks with tool access and real-world consequences.

Question 8

Question Type: MultipleChoice

A security analyst notices that regardless of user-submitted prompts, an AI model always returns unsanitized responses. These responses are then passed to multiple plug-ins. The analyst is concerned with the potential security implications.

Which option best Open Worldwide Application Security Project (OWASP) categories addresses this vulnerability?

Options:

A- Misinformation
B- Prompt injection
C- Unbounded consumption
D- Improper output handling

Answer:

D

Explanation:

Basic Concept: OWASP has published the Top 10 vulnerabilities for Large Language Model Applications, each addressing a distinct category of LLM security risk. Understanding which OWASP category maps to specific LLM vulnerability scenarios is a key competency in the CompTIA SecAI+ Study Guide under securing AI systems.

Why D is Correct: Improper output handling (OWASP LLM02) occurs when an application passes LLM-generated outputs to downstream systems such as plug-ins, web browsers, or databases without proper validation, sanitization, or encoding. This can enable XSS, SQL injection, remote code execution, or other injection attacks against plug-ins and downstream

systems. The scenario exactly matches this: unsanitized AI responses are automatically passed to multiple plug-ins, which could execute malicious content in the model's output.

Why A is Wrong: Misinformation refers to the AI generating false or misleading content that users might believe. It is a content accuracy concern related to hallucinations and false information propagation, not a vulnerability describing how model outputs are handled by downstream systems.

Why B is Wrong: Prompt injection involves crafting inputs to manipulate model behavior and override instructions. While it can be a contributing cause of unsafe outputs, the vulnerability described --- passing unsanitized outputs to plug-ins --- is specifically the output handling failure, not the injection mechanism itself.

Why C is Wrong: Unbounded consumption (OWASP LLM10) refers to resource exhaustion attacks including denial-of-wallet and denial-of-service through excessive token consumption. It addresses resource management vulnerabilities, not the security implications of passing model outputs to downstream systems.

Question 9

Question Type: MultipleChoice

An administrator must conduct generative AI cost monitoring for use in the healthcare industry.

Which of the following criteria is the best way to calculate this cost?

Options:

A- Connection access and exchange gateway
B- Encryption and decryption processing
C- Storage retrieval and prompt processing
D- Catalog servicing and exchange processing

Answer:

C

Explanation:

Basic Concept: Generative AI systems in healthcare settings incur costs from multiple operational activities. Understanding the cost drivers specific to generative AI helps administrators implementaccurate cost monitoring and controls. CompTIA SecAI+ Study Guide covers AI cost management under securing AI systems.

Why C is Correct: Storage retrieval and prompt processing are the two primary cost drivers for generative AI systems in healthcare. Storage retrieval refers to the cost of querying vector databases or document stores in RAG-based AI systems to fetch relevant patient records, clinical guidelines, or historical data for context. Prompt processing encompasses the token-based cost of the LLM processing the combined retrieved content and user query to generate a response. Together these two activities represent the billable units that drive generative AI costs in healthcare RAG deployments, making them the most accurate basis for cost calculation and monitoring.

Why A is Wrong: Connection access and exchange gateway costs relate to network infrastructure and API gateway usage fees. While there may be minor costs associated with API calls, these are not the primary cost drivers for generative AI systems where the dominant expenses are computational token processing and data retrieval operations.

Why B is Wrong: Encryption and decryption processing costs relate to cryptographic operations for data security. While encryption is important for healthcare data protection under HIPAA, cryptographic processing overhead is minimal compared to the substantial token-based LLM processing and storage retrieval costs that dominate generative AI operational expenses.

Why D is Wrong: Catalog servicing and exchange processing are terms associated with data catalog management and data exchange infrastructure. These are not recognized primary cost components of generative AI systems in healthcare, where storage retrieval and token-based prompt processing are the established cost measurement criteria.

Question 10

Question Type: MultipleChoice

An organization recently developed an AI-powered product and discovers that it is vulnerable to attacks in which malicious actors can alter the input, causing the system to recommend inappropriate information.

Which of the following techniques is the most effective way to secure the system against manipulation attacks?

Options:

A- Cross-validation
B- Feature regularization
C- Feature scaling
D- Guardrails

Answer:

D

Explanation:

Basic Concept: Input manipulation attacks --- including adversarial examples and prompt injection --- alter inputs to cause AI systems to produce unintended, harmful, or inappropriate outputs. Defending against these attacks requires mechanisms that validate and constrain both inputs and outputs at runtime. CompTIA SecAI+ Study Guide identifies guardrails as the primary defense against input manipulation in AI systems.

Why D is Correct: Guardrails implement real-time validation and filtering of both incoming inputs and outgoing recommendations. They detect manipulated inputs that deviate from expected patterns, enforce content policies on outputs, and prevent the system from producing inappropriate recommendations regardless of how cleverly the input was crafted. Guardrails provide the most comprehensive and directly applicable defense against the described manipulation attack scenario.

Why A is Wrong: Cross-validation is a model evaluation technique that assesses how well a model generalizes to independent datasets during training. It measures predictive performance but does not provide runtime protection against input manipulation attacks on a deployed system.

Why B is Wrong: Feature regularization is a training technique that adds penalties to model weights to prevent overfitting. It improves generalization during training but does not inspect or validate inputs at inference time to detect manipulation.

Why C is Wrong: Feature scaling normalizes input feature values to a standard range for training efficiency. Like regularization, it is a preprocessing and training step that has no effect on defending against runtime input manipulation attacks on deployed systems.

Question 11

Question Type: MultipleChoice

A data scientist is working with unlabeled data and wants to build a clustering model.

Which of the following techniques should a data scientist use?

Options:

A- Supervised learning
B- Reinforcement learning
C- Unsupervised learning
D- Semi-supervised learning

Answer:

C

Explanation:

Basic Concept: Different ML learning paradigms handle different data situations. The availability of labeled versus unlabeled data determines which learning approach is appropriate. Building clustering models specifically requires learning from data without predefined category labels. CompTIA SecAI+ Study Guide covers ML learning paradigms under basic AI concepts.

Why C is Correct: Unsupervised learning works with unlabeled data by discovering inherent patterns, structures, and groupings within the data without predefined categories. Clustering is the canonical unsupervised learning task, where algorithms like k-means, hierarchical clustering, or DBSCAN group similar data points together based on feature similarity. Since the data scientist has unlabeled data and wants to find natural groupings, unsupervised learning is the appropriate and correct technique.

Why A is Wrong: Supervised learning requires labeled training data where each example has a corresponding correct output label. The data scientist explicitly has unlabeled data, making supervised learning inapplicable without first completing the labor-intensive task of manually labeling all examples.

Why B is Wrong: Reinforcement learning trains agents to take actions in an environment to maximize cumulative rewards through trial and error. It is designed for sequential decision-making problems, not for finding groupings in static, unlabeled datasets.

Why D is Wrong: Semi-supervised learning combines a small amount of labeled data with a large amount of unlabeled data. It requires at least some labels to guide learning. The scenario specifies working with unlabeled data only, making unsupervised learning the pure fit.

Question 12

Question Type: MultipleChoice

During a model validation procedure, an engineer notices that a model performs well during training but poorly during testing.

Which option best best describes the reason?

Options:

A- Fine-tuning
B- Overfitting

C- Regularization
D- Inference

Answer:

B

Explanation:

Basic Concept: The gap between training performance and test performance is a classic indicator of a specific model quality problem. Understanding this phenomenon and its causes is fundamental to AI model development. CompTIA SecAI+ Study Guide covers overfitting under basic AI concepts and model quality.

Why B is Correct: Overfitting occurs when a model learns the training data too specifically --- memorizing noise, outliers, and specific patterns in the training set rather than learning generalizable underlying patterns. The model achieves high accuracy on training data but fails to generalize to new, unseen test data. This produces exactly the scenario described: excellent training performance combined with poor test performance. Overfitting is the quintessential cause of this training-testing performance gap.

Why A is Wrong: Fine-tuning is a training technique that adapts a pre-trained model to a new task or domain using additional training data. It is a deliberate training process, not a description of why a model's performance degrades from training to testing.

Why C is Wrong: Regularization is a training technique specifically used to prevent overfitting by adding penalties to large model weights, encouraging the model to learn simpler, more generalizable patterns. It is the solution to overfitting, not its cause.

Why D is Wrong: Inference is the process of using a trained model to make predictions on new data. It describes the operational use of a model, not a quality characteristic that explains why performance differs between training and testing phases.