CERT EMPIRE

ISACA AAISM Exam Questions

Total Questions: 80+
Demo Questions: 15
Version: Updated for 2025

Prepared and Verified by Cert Empire – Your Trusted IT Certification Partner

For Access to the full set of Updated Questions – Visit:
ISACA AAISM Exam Questions by Cert Empire

Question: 1

When implementing a generative AI system, which of the following approaches will BEST prevent misalignment between the corporate risk appetite and tolerance?

A. Ensuring effective AI key performance indicators (KPIs)

B. Performing an AI impact assessment

C. Creating and maintaining an AI risk register

D. Establishing and monitoring acceptable levels of AI system risk

Answer:

D

Explanation:

Establishing and monitoring acceptable levels of AI system risk is the most direct and effective approach. This process translates the organization's high-level risk appetite and tolerance into specific, measurable, and actionable thresholds for the AI system. "Establishing" sets the boundaries for what is acceptable, directly aligning the system with corporate policy. "Monitoring" provides the ongoing governance mechanism to ensure the system operates within these predefined boundaries, actively preventing deviations and thus ensuring continuous alignment. This creates a complete feedback loop for risk management.

Why Incorrect Options are Wrong:

A. KPIs are metrics used for monitoring performance; while they can support risk management, they do not, by themselves, establish the acceptable risk framework or ensure alignment.
B. An AI impact assessment is a critical, but typically point-in-time, activity to identify potential risks. It informs the risk strategy but does not constitute the ongoing alignment process.
C. An AI risk register is a tool for documenting and tracking identified risks. It is a static record and a component of risk management, not the active process of setting and enforcing limits.

References:

1. ISACA, Artificial Intelligence Audit Framework, 2023. In the "AI Governance" domain, the framework emphasizes that governance structures must "define and approve the risk appetite and tolerance for AI initiatives" and ensure that "AI risk management is integrated with the enterprise risk management framework." Establishing and monitoring acceptable levels is the operationalization of this principle. (Domain 1, AI Governance).
2. National Institute of Standards and Technology (NIST), AI Risk Management Framework (AI RMF 1.0), January 2023. The GOVERN function is foundational and states a key outcome is that "The organization's risk management processes are informed by its understanding of AI system-related risk and its established risk tolerance." This directly supports establishing risk

levels based on tolerance and then using other functions (MAP, MEASURE, MANAGE) to monitor and control them. (Section 3.1, GOVERN Function).

3. Deloitte, AI risk management framework: A structured approach to managing AI risks, 2023. This publication, aligned with ISACA and NIST principles, notes that a key component of an AI risk framework is to "Define risk appetite and tolerance levels" and then "Establish key risk indicators (KRIs) and thresholds to monitor AI risks against defined tolerance levels." This highlights the direct link between setting levels and monitoring them to prevent misalignment. (Page 8, "Define risk appetite and tolerance levels").

Question: 2

Which of the following BEST reduces the risk of exposing sensitive data through the output of large language models (LLMs) in applications?

A. Encrypting data in transit and at rest

B. Conducting adversarial testing

C. Implementing data sanitization techniques

D. Enforcing least privilege access

Answer:

C

Explanation:

The most effective method to prevent large language models (LLMs) from exposing sensitive data in their output is to ensure they are never exposed to it in the first place. Data sanitization techniques, such as masking, redaction, or tokenization, are applied to the training data and user prompts to remove or anonymize personally identifiable information (PII) and other sensitive data before the LLM processes it. By cleaning the input, the model does not learn the sensitive information and therefore cannot inadvertently reproduce or "regurgitate" it in its responses. This is a direct, preventative control that addresses the root cause of the risk.

Why Incorrect Options are Wrong:

A. Encrypting data in transit and at rest protects data during storage and transmission but not during processing. The LLM requires decrypted data to function, at which point it can be exposed.
B. Conducting adversarial testing is a validation technique used to identify if an LLM is vulnerable to leaking sensitive data, rather than a preventative control that reduces the risk itself.
D. Enforcing least privilege access restricts who can interact with the LLM or its underlying data, but it does not prevent an authorized user or process from using sensitive data that the model can then leak.

References:

1. ISACA, Artificial Intelligence Audit Toolkit, 2023. In the "AI Governance and Risk Management" domain, Control Objective A&G-05: "Data Privacy and Protection" emphasizes implementing controls to protect sensitive data throughout the AI lifecycle. This includes the data preprocessing stage, where sanitization is a key activity to prevent PII from being ingested by the model.
2. ISACA, Auditing Artificial Intelligence White Paper, 2023. Page 12, under the "Data Acquisition and Preparation" section, states that a key audit step is to "verify that controls are in place to ensure that PII and other sensitive data are removed or anonymized from data sets prior to their use in AI/ML models." This directly supports sanitization as the primary control.

3. Mee, T., et al. A Survey on Privacy in Large Language Models: Attacks, Defenses, and Future Directions. arXiv, 2023. Section 4.1, "Data Pre-processing," identifies data sanitization and anonymization as primary defense mechanisms against privacy attacks. It states, "The most straightforward defense is to remove sensitive information from the pre-training data." (DOI: https://doi.org/10.48550/arXiv.2312.06717).

Question: 3

Which of the following is the GREATEST risk inherent to implementing generative AI?

A. Lack of employee training

B. Unidentified asset vulnerabilities

C. Inadequate return on investment (ROI)

D. Potential intellectual property violations

Answer:

D

Explanation:

The greatest inherent risk in implementing generative AI is the potential for intellectual property (IP) violations. This risk is twofold and fundamental to how these models operate. First, generative AI models are trained on vast datasets, which often include copyrighted materials without permission, creating a risk that the model's output could be considered a derivative work and thus an infringement. Second, employees may input sensitive or proprietary company data (trade secrets, confidential information) into public AI tools, which can lead to the data being absorbed by the model and potentially exposed, resulting in a catastrophic loss of the organization's own IP.

Why Incorrect Options are Wrong:

A. Lack of employee training: This is a significant operational risk for any new technology implementation, not a risk uniquely inherent to the core function of generative AI itself.
B. Unidentified asset vulnerabilities: This is a general cybersecurity risk applicable to all software and IT systems, rather than the most significant risk specific to the generative nature of AI.
C. Inadequate return on investment (ROI): This is a financial or business risk common to any major project or investment and is not a technical or legal risk inherent to the AI technology.

References:

1. ISACA. (2023). Generative AI Governance and Risk Management. Page 6, Section "Key Risks of Generative AI." This guide explicitly identifies "Intellectual Property (IP) and Copyright Infringement" as a primary risk, stating, "Generative AI models are trained on vast amounts of data, which may include copyrighted material... This can lead to legal challenges and potential liability for copyright infringement if the generated content is substantially similar to the original work."
2. ISACA. (2023). Artificial Intelligence: An Audit and Assurance Framework. Page 23, Section 3.2 "AI Risk Universe." The framework details data-related risks, including the use of proprietary or copyrighted data for training AI models, which can result in "legal, regulatory, reputational and

financial impacts."
3. Stanford Institute for Human-Centered Artificial Intelligence (HAI). (2023). A Guide to Generative AI. Section "Risks and Challenges." The guide discusses IP as a major challenge, noting the legal ambiguity and risk surrounding the use of copyrighted data for training models and the potential for generated outputs to infringe on existing copyrights.

Question: 4

Which of the following should be the PRIMARY consideration for an organization concerned about liabilities associated with unforeseen behavior from agentic AI systems?

A. Model dependencies

B. Approved base models

C. Accountability model

D. Acceptable risk level

Answer:

C

Explanation:

The primary consideration for managing liabilities from agentic AI's unforeseen behavior is establishing a robust accountability model. This model defines and assigns responsibility for the AI's decisions and outcomes throughout its lifecycle. When an autonomous system acts in an unexpected way, a clear accountability framework is essential for determining legal and ethical responsibility, guiding incident response, and addressing potential harm. It provides the foundational governance structure for managing the specific risk of liability.

Why Incorrect Options are Wrong:

A. Model dependencies are a technical aspect of an AI system's architecture; while important for operational risk, they do not primarily address the legal framework for liability.
B. Using approved base models is a risk mitigation control, but it does not eliminate the possibility of unforeseen behavior or define who is liable when it occurs.
D. An acceptable risk level is a strategic business decision that sets the threshold for risk tolerance, but it is not the mechanism for managing liability when that threshold is breached.

References:

1. ISACA, Auditing Artificial Intelligence, 2023: Chapter 2, "AI Governance," emphasizes that a key principle of AI governance is accountability. It states, "Accountability for AI systems should be established, with clear roles and responsibilities for the development, deployment, and ongoing monitoring of AI systems." This directly links accountability models to the management of AI systems and their outcomes.
2. National Institute of Standards and Technology (NIST), AI Risk Management Framework (AI RMF 1.0), January 2023: The "Govern" function is foundational to the framework. Section 3.1, GOVERN-1, highlights the need to establish "policies, processes, procedures, and practices for AI risk management, including roles, responsibilities, authorities, and lines of communication." This establishes an accountability structure as a primary step in managing AI risk, including

liabilities.

3. Stanford University Human-Centered AI (HAI), An Action Plan for Advancing Trustworthy AI, 2023: The report repeatedly stresses the importance of accountability mechanisms. On page 10, it calls for "clear accountability mechanisms to ensure that there is recourse and redress when AI systems cause harm," identifying this as a critical component for building trustworthy AI and managing its societal impact, including liability.

Question: 5

Which of the following is the GREATEST benefit of implementing an AI tool to safeguard sensitive data and prevent unauthorized access?

A. Timely analysis of endpoint activities

B. Timely initiation of incident response

C. Reduced number of false positives

D. Reduced need for data classification

Answer:

C

Explanation:

The greatest benefit of implementing AI in this context is its ability to significantly reduce the number of false positives. AI and machine learning models excel at establishing a highly accurate baseline of normal user and system behavior. By understanding what is normal, they can more precisely identify true anomalies indicative of unauthorized access or data exfiltration. This capability directly addresses the critical operational challenge of "alert fatigue," where security analysts are overwhelmed by alerts from traditional rule-based systems. By reducing this noise, AI allows security teams to focus their time and resources on investigating and responding to credible, high-priority threats.

Why Incorrect Options are Wrong:

A. Timely analysis of endpoint activities: While AI enables rapid analysis, the value of this speed is diminished if the output is filled with false alarms, making it difficult to discern genuine threats.
B. Timely initiation of incident response: This is a consequence of accurate detection. Initiating responses to a high volume of false positives can be disruptive and counterproductive, making this a secondary benefit to accuracy.
D. Reduced need for data classification: This is incorrect. AI security tools rely heavily on proper data classification to function effectively. The AI must know which data is sensitive to apply appropriate protective measures.

References:

1. ISACA. (2021). Using AI to Improve Cybersecurity. ISACA Journal, Volume 4. Retrieved from ISACA official publications. The article notes, "AI can help reduce the number of false positives... By learning what normal network traffic looks like, AI can more accurately identify suspicious activity... This allows security teams to focus on real threats and respond more quickly and effectively."
2. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for

cyber security. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176. (Section IV.A, Anomaly Detection). This survey highlights that a major challenge for traditional intrusion detection systems is the "high rate of false alarms," a problem that advanced machine learning methods are designed to mitigate by improving detection accuracy.
https://doi.org/10.1109/COMST.2015.2494502
3. ISACA. (2023). Auditing Artificial Intelligence. In the context of AI applications in security (Chapter 4: AI Use Cases), the guide discusses how AI-powered systems improve upon legacy systems by providing higher-fidelity alerts, which is a direct reference to reducing false positives and increasing the accuracy of threat detection.

Question: 6

Which of the following employee awareness topics would MOST likely be revised to account for AI- enabled cyber risk?

A. Clean desk policy

B. Social engineering

C. Malicious insider threats

D. Authentication controls

Answer:

B

Explanation:

Artificial intelligence, particularly generative AI, has dramatically increased the sophistication and scalability of social engineering attacks. AI can create highly convincing, context-aware phishing emails, text messages, and even deepfake audio and video for vishing (voice phishing) attacks. These AI-generated attacks can bypass traditional human detection methods that rely on spotting poor grammar or generic messaging. Therefore, employee awareness training must be significantly revised to educate staff on these new, hyper-realistic threats, teaching them to verify requests through out-of-band channels and to be skeptical of even seemingly legitimate voice or video communications.

Why Incorrect Options are Wrong:

A. Clean desk policy: This policy's core principles-securing physical documents and locking screens-are not fundamentally changed by AI-enabled cyber risks.
C. Malicious insider threats: While an insider might use AI as a tool, the fundamental awareness training for this threat, which focuses on behavioral red flags and access control principles, remains largely the same.
D. Authentication controls: AI primarily impacts the technical side of authentication (e.g., credential stuffing attacks). Employee awareness training continues to focus on established best practices like using strong passwords and multi-factor authentication (MFA).

References:

1. ISACA Journal, Volume 1, 2024, "AI-Enabled Cyberattacks": This article discusses how AI is a "game changer for social engineering," noting that generative AI can "create phishing emails that are grammatically perfect, contextually relevant and highly persuasive," thus requiring new forms of employee training and awareness.
2. NIST, "AI Risk Management Framework (AI RMF 1.0)," January 2023, Section 3.3, "AI Risk and Trustworthiness": The framework discusses risks associated with malicious AI use. It

implicitly supports the need for updated security awareness by highlighting AI's capability to "generate deceptive or manipulative content" (p. 13), which is the foundation of modern social engineering attacks.

3. Seymour, J., & Tully, J. (2017). "Weaponized AI: The new face of social engineering." Black Hat USA Conference Proceedings.: This academic conference paper details how AI can be used to automate and personalize social engineering attacks at a massive scale, making traditional awareness training less effective and necessitating a revised approach that accounts for AI-driven tactics.

4. MIT Technology Review, "The biggest security threats of 2024 are all about AI," January 2, 2024: This article highlights that security experts are most concerned about AI-powered phishing and disinformation campaigns. It explains that generative AI makes it "easier and cheaper to run social engineering campaigns," directly impacting what employees need to be aware of.

Question: 7

Which of the following BEST enables an organization to maintain visibility to its AI usage?

A. Ensuring the board approves the policies and standards that define corporate AI strategy

B. Maintaining a monthly dashboard that captures all AI vendors

C. Maintaining a comprehensive inventory of AI systems and business units that leverage them

D. Measuring the impact of AI implementation using key performance indicators (KPIs)

Answer:

C

Explanation:

A comprehensive inventory is the most fundamental and direct mechanism for maintaining visibility into an organization's AI usage. It serves as a central repository that documents all AI systems, models, and applications, whether developed in-house or procured from vendors. By linking these systems to the specific business units that leverage them, the organization gains a clear, enterprise-wide view of its AI footprint. This inventory is the foundational element for effective AI governance, risk management, and strategic oversight, directly enabling continuous visibility.

Why Incorrect Options are Wrong:

A. Board approval of policies establishes the high-level governance framework but does not provide the operational, ongoing visibility into specific AI systems being used.
B. A vendor dashboard is incomplete as it overlooks internally developed AI systems and does not provide the necessary detail on how specific applications are being used.
D. Measuring impact with KPIs is a post-implementation activity focused on performance and value. It relies on first having visibility, which the inventory provides.

References:

1. ISACA, Artificial Intelligence Audit Framework, 2023. In the "AI Governance" domain, Control Objective GOV-02, "AI Inventory Management," states the need to "Establish and maintain a comprehensive inventory of all AI systems used within the organization to ensure proper oversight and management." This directly supports the inventory as the key to visibility.
2. ISACA, Auditing Artificial Intelligence, 2021. Page 13, under the section "Develop an AI Audit Plan," specifies, "The first step in developing an AI audit plan is to create an inventory of AI use cases... The inventory should be a living document that is updated as new AI use cases are identified." This highlights the inventory as the primary tool for awareness and visibility.
3. Kozyrkov, C. (2020). AI Governance: A Primer for Boards of Directors. Stanford University Human-Centered AI Institute (HAI). This publication, while aimed at boards, implicitly supports the

need for inventories by discussing the board's responsibility for overseeing AI risks. Effective oversight is impossible without a clear inventory of what AI systems the organization possesses. The concept is foundational to the "Know Your AI" principle of governance.

Question: 8

Which of the following is the MOST important course of action prior to placing an in-house developed AI solution into production?

A. Perform a privacy, security, and compliance gap analysis

B. Deploy a prototype of the solution

C. Obtain senior management sign-off

D. Perform testing, evaluation, validation, and verification

Answer:

D

Explanation:

Performing comprehensive Testing, Evaluation, Validation, and Verification (TEVV) is the most critical technical prerequisite before deploying an AI solution. This process ensures the system meets its specified requirements for functionality, performance, reliability, security, and fairness. TEVV provides the objective evidence needed to confirm that the AI model behaves as intended in the target operational environment and that associated risks are identified and mitigated. Without successful TEVV, there is no assurance that the system is fit for purpose, making deployment irresponsible and exposing the organization to significant operational, financial, and reputational risks.

Why Incorrect Options are Wrong:

A. This analysis is a crucial activity, but it should be performed iteratively throughout the AI lifecycle, not just as a final pre-deployment step.

B. A prototype is an early-stage model used for proof-of-concept and feasibility studies; it is not the version that would be placed into production.

C. Senior management sign-off is a critical governance gate, but this approval is fundamentally dependent on the successful results and evidence produced by the TEVV process.

References:

1. ISACA, Artificial Intelligence Audit Framework, 2023: Domain 4, "AI Model Development and Implementation," Control Objective AI.4.5 "Testing and Validation," states, "Ensure that the AI model undergoes rigorous testing and validation to verify its performance, accuracy and reliability before deployment." This highlights TEVV as the essential pre-deployment verification step.
2. National Institute of Standards and Technology (NIST), AI Risk Management Framework (AI RMF 1.0), January 2023: The "Measure" function (Section 3.3, page 17) is dedicated to activities that assess AI risks. It explicitly includes "Testing, Evaluation, Validation, and Verification (TEVV)" as a core category (MEASURE 1), emphasizing that these evaluations are necessary to make

informed decisions about AI system deployment and to ensure it functions as intended.
3. Stanford University, CS 329S: Machine Learning Systems Design, Winter 2021 Lecture 8 "MLOps & Tooling": The courseware outlines the ML project lifecycle, where rigorous testing and evaluation are depicted as the final technical stage before a model is "pushed to production." This confirms that comprehensive testing is the immediate precursor to deployment in established MLOps practices.

Question: 9

An organization decides to contract a vendor to implement a new set of AI libraries. Which of the following is MOST important to address in the master service agreement to protect data used during the AI training process?

A. Data pseudonymization

B. Continuous data monitoring

C. Independent certification

D. Right to audit

Answer:

D

Explanation:

The right to audit is a contractual clause in the master service agreement (MSA) that grants an organization the legal authority to inspect and verify a vendor's controls, processes, and adherence to security requirements. When dealing with sensitive AI training data, this right is paramount. It provides the ultimate mechanism for assurance, allowing the organization to directly confirm that all other specified protections (such as pseudonymization, monitoring, and data handling policies) are being implemented effectively. It is the most fundamental contractual tool for maintaining oversight and managing third-party risk.

Why Incorrect Options are Wrong:

A. Data pseudonymization: This is a specific technical data protection technique. While important, the right to audit is the contractual mechanism needed to verify that pseudonymization is actually being performed correctly.

B. Continuous data monitoring: This is an operational security control. The right to audit provides the means to ensure that this monitoring is in place, is effective, and meets contractual requirements.

C. Independent certification: While valuable, a certification (e.g., SOC 2, ISO 27001) provides point-in-time assurance and may not cover the specific scope of the AI implementation or the organization's unique data.

References:

1. ISACA, Auditing Artificial Intelligence, 2021: This official ISACA publication states, "Contracts with third-party providers should include clauses that allow for the auditing of the AI system, including its algorithms, data and controls. This is especially important when the AI system is used for critical functions or processes." (Page 19, Section: "Third-party AI Systems"). This directly supports the necessity of audit rights in vendor agreements for AI systems.

2. ISACA, Artificial Intelligence Audit Toolkit, 2023: In the "AI Governance and Risk Management" domain, Program Step 1.4, "Evaluate Vendor Management," emphasizes reviewing contracts for key provisions. The ability to assess vendor compliance, which is enabled by a right-to-audit clause, is a core component of this evaluation. The toolkit's focus is on verifiable controls, and the right to audit is the primary contractual method for such verification.
3. Tsamados, A., et al. (2022). The ethics of algorithms: key problems and solutions. The Alan Turing Institute. While discussing AI governance and accountability, the paper highlights the need for "mechanisms for verification and audit" when relying on third-party systems. This academic consensus underscores that contractual audit rights are essential for external accountability. (Section 4.3, "Accountability"). DOI: https://doi.org/10.1080/25741292.2021.1976502

Question: 10

Which of the following is the MOST effective use of AI in incident response?

A. Streamlining incident response testing

B. Automating incident response triage

C. Improving incident response playbook

D. Ensuring chain of custody

Answer:

B

Explanation:

The most effective use of AI in incident response is automating the triage process. Incident triage involves sorting, prioritizing, and assigning the vast number of alerts generated by security tools. This is a time-consuming, repetitive, and data-intensive task for human analysts, often leading to "alert fatigue." AI, particularly machine learning, excels at rapidly analyzing large datasets, identifying patterns, correlating events, and classifying alerts with high accuracy. By automating triage, organizations can significantly reduce the Mean Time to Detect (MTTD) and Mean Time to Respond (MTTR), allowing security teams to focus their expertise on investigating and resolving the most critical incidents, thereby minimizing the potential impact of an attack.

Why Incorrect Options are Wrong:

A. Streamlining incident response testing: While AI can be used to create more sophisticated attack simulations for testing, its impact on real-time operational efficiency is less direct than its role in triage.

C. Improving incident response playbook: AI can analyze past incidents to suggest playbook improvements, but this is a strategic, post-incident activity, not a direct application that enhances the immediate response to an ongoing threat.

D. Ensuring chain of custody: Chain of custody is a critical forensic and procedural process. While AI can assist in logging and tracking digital evidence, ensuring its integrity is primarily reliant on cryptographic hashing and strict procedural controls, not AI-driven decision-making.

References:

1. ISACA White Paper, Artificial Intelligence for a More Resilient Enterprise, 2021: This publication states, "AI can automate the initial triage of security alerts, freeing up security analysts to focus on more complex threats. This can help to reduce the time it takes to detect and respond to incidents, and it can also improve the accuracy of incident response." (Section: "AI for Cybersecurity," Paragraph 3).
2. NIST Special Publication 800-61 Rev. 2, Computer Security Incident Handling Guide, 2012:

While published before the widespread adoption of AI, this foundational document emphasizes the importance of timely and accurate analysis in the "Detection and Analysis" phase (Section 3.2.2). Modern AI-driven Security Orchestration, Automation, and Response (SOAR) platforms directly address this need for speed and accuracy in triage, which is a core part of this phase.
3. Ullah, I., & Mahmoud, Q. H. (2022). A Comprehensive Survey on the Use of Artificial Intelligence in Cybersecurity: A Scoping Review. IEEE Access, 10, 55731-55754. https://doi.org/10.1109/ACCESS.2022.3177139: This peer-reviewed survey highlights that a primary application of AI in cybersecurity is to "handle the overwhelming number of alerts generated by security systems" by "automating the process of alert triage and prioritization." (Section IV.A, "Threat Detection and Incident Response").

Question: 11

An automotive manufacturer uses AI-enabled sensors on machinery to monitor variables such as vibration, temperature, and pressure. Which of the following BEST demonstrates how this approach contributes to operational resilience?

A. Scheduling repairs for critical equipment based on real-time condition monitoring

B. Performing regular maintenance based on manufacturer recommendations

C. Conducting monthly manual reviews of maintenance schedules

D. Automating equipment repairs without any human intervention

Answer:

A

Explanation:

The use of AI-enabled sensors for real-time condition monitoring is a core component of predictive maintenance (PdM). By continuously analyzing operational data such as vibration, temperature, and pressure, the AI system can identify patterns that precede equipment failure. This allows the organization to schedule repairs proactively, just before a fault is likely to occur, thereby preventing unexpected breakdowns and minimizing unplanned downtime. This direct avoidance of operational disruption is a primary contributor to enhancing operational resilience in a manufacturing environment.

Why Incorrect Options are Wrong:

B. This describes a traditional, preventative (time-based) maintenance schedule, which does not leverage the real-time, condition-based data from the AI sensors mentioned.
C. This is a manual, administrative task. It is a reactive or periodic review rather than a dynamic, data-driven action enabled by the AI system.
D. The scenario describes AI for monitoring and data analysis, not for the physical execution of automated repairs, which is a different and more advanced capability.

References:

1. ISACA. (2021). Auditing Artificial Intelligence White Paper. Page 8. The paper notes that AI applications can lead to "improved operational efficiency and reduced downtime," which is achieved through capabilities like predictive maintenance, directly supporting the concept of operational resilience.
2. Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, 106889. Section 3.1 discusses how AI and machine learning models use sensor data to predict the "Remaining Useful Life (RUL)" of equipment, enabling maintenance

to be scheduled to prevent failures. https://doi.org/10.1016/j.cie.2020.106889
3. Massachusetts Institute of Technology (MIT) OpenCourseWare. (2015). 2.830J Control of Manufacturing Processes (SMA 6303). Lecture 1: Introduction to Manufacturing Process Control. The course materials explain the principle of using in-process sensing to monitor process variables to detect and prevent deviations that could lead to defects or equipment failure, which is the foundational concept behind the scenario.

Question: 12

Which of the following BEST describes how supervised learning models help reduce false positives in cybersecurity threat detection?

A. They analyze patterns in data to group legitimate activity from actual threats

B. They use real-time feature engineering to automatically adjust decision boundaries

C. They learn from historical labeled data

D. They dynamically generate new labeled data sets

Answer:

C

Explanation:

Supervised learning is a machine learning paradigm where an algorithm learns from a dataset that has been manually labeled with the correct outcomes. In cybersecurity, this involves training a model on historical data where events are explicitly tagged as either "malicious" or "benign." The model learns the patterns and features that distinguish these two classes. By training on a high-quality, well-labeled dataset that accurately represents both legitimate and threatening activities, the model can build a robust decision boundary. This allows it to more accurately classify new, unseen data, thereby reducing the number of times it incorrectly flags legitimate activity as a threat (a false positive).

Why Incorrect Options are Wrong:

A. This describes unsupervised learning (e.g., clustering), which finds inherent patterns to group data without relying on pre-existing labels.
B. While some advanced models can adjust in real-time (online learning), the fundamental principle of supervised learning is training on a static, historical labeled dataset.
D. This describes techniques like data augmentation or synthetic data generation, which are used to supplement a training set, not the core learning mechanism itself.

References:

1. ISACA. (2021). Artificial Intelligence for Auditing. "Supervised learning uses labeled data sets to train algorithms to classify data or predict outcomes accurately. With supervised learning, the enterprise provides the AI model with both inputs and desired outputs." (Page 8, "Supervised Learning" section).
2. Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., ... & Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. IEEE Access, 6, 35365-35381. "In supervised learning, the training data consist of a set of training examples, where each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal)."

(Section II-A, Paragraph 1). https://doi.org/10.1109/ACCESS.2018.2837699
3. Ng, A. (2008). CS229 Machine Learning Course Notes. Stanford University. "In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output." (Part I, "Supervised Learning," Page 2).

Question: 13

Which of the following BEST represents a combination of quantitative and qualitative metrics that can be used to comprehensively evaluate AI transparency?

A. AI system availability and downtime metrics

B. AI model complexity and accuracy metrics

C. AI explainability reports and bias metrics

D. AI ethical impact and user feedback metrics

Answer:

C

Explanation:

AI transparency is evaluated by understanding a model's internal logic and its fairness. This requires a mix of metric types. Explainability reports offer qualitative, human-interpretable narratives about how a model arrives at its decisions, directly addressing the "black box" problem. Bias metrics (e.g., disparate impact, equal opportunity difference) provide quantitative, statistical evidence of whether the model produces systematically unfair outcomes for different demographic groups. This combination of qualitative explanations and quantitative fairness measurements provides a direct and comprehensive assessment of an AI system's transparency, which is fundamental for accountability and trust.

Why Incorrect Options are Wrong:

A. AI system availability and downtime metrics
These are purely quantitative operational metrics that measure system reliability, not its transparency or decision-making logic.
B. AI model complexity and accuracy metrics
These are primarily quantitative performance and structural metrics. While complexity can inversely relate to interpretability, they do not offer a comprehensive view of transparency.
D. AI ethical impact and user feedback metrics
These are broader measures. Ethical impact is a high-level qualitative assessment, while user feedback measures perception rather than the system's intrinsic transparent properties.

References:

1. National Institute of Standards and Technology (NIST). (2023). AI Risk Management Framework (AI RMF 1.0).
Section 4.2.2, "MEASURE," discusses the need to identify metrics and methodologies to assess AI risks, including those related to bias and interpretability/explainability. It states, "Metrics may be qualitative or quantitative" (p. 24).

Section 3.3, "Characteristics of Trustworthy AI," defines transparency as including explainability and interpretability, which involves providing access to information about how an AI system works (p. 14).

2. ISACA. (2023). Auditing Artificial Intelligence.

Chapter 3, "AI Risks and Controls," explicitly links transparency to explainability, stating, "Transparency is the extent to which the inner workings of an AI system are understandable to humans... Explainable AI (XAI) is a set of techniques and methods that help to make AI systems more transparent" (p. 31). The chapter also details the risk of bias and the need for metrics to detect it (p. 33).

3. Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. Information Fusion, 58, 82-115.

Section 2, "The pillars of responsible AI," identifies transparency as a key pillar, which is achieved through explainability (qualitative descriptions) and the assessment of fairness and bias (often using quantitative metrics). DOI: https://doi.org/10.1016/j.inffus.2019.12.012

Question: 14

Which of the following key risk indicators (KRIs) is MOST relevant when evaluating the effectiveness of an organization's AI risk management program?

A. Number of AI models deployed into production

B. Percentage of critical business systems with AI components

C. Percentage of AI projects in compliance

D. Number of AI-related training requests submitted

Answer:

C

Explanation:

A Key Risk Indicator (KRI) for an AI risk management program should measure the program's effectiveness in governing AI initiatives and ensuring they adhere to established policies and controls. The "Percentage of AI projects in compliance" is the most direct measure of this effectiveness. It quantifies how well the organization's AI activities are following the prescribed risk management framework, including mandatory assessments, controls, and documentation. A high compliance rate indicates a successful and effective program, while a low rate serves as an early warning that the program is not being implemented properly, increasing overall AI-related risk.

Why Incorrect Options are Wrong:

A. Number of AI models deployed into production: This is a volume or activity metric. It indicates the scale of AI adoption and potential risk exposure but does not measure the effectiveness of the program managing that risk.
B. Percentage of critical business systems with AI components: This metric measures the organization's inherent risk or attack surface related to AI. It identifies where risk management is crucial but does not evaluate how well it is being performed.
D. Number of AI-related training requests submitted: This is an ambiguous indicator. It could signify a positive culture of risk awareness or, conversely, a lack of adequate foundational training, but it does not directly measure the program's control effectiveness.

References:

1. NIST AI Risk Management Framework (AI RMF 1.0): The "Measure" function of the framework is dedicated to tracking risk management effectiveness. It states, "Measurement enables learning from experience and improves the design, development, deployment, and use of AI systems." A compliance metric directly aligns with this goal of evaluating and improving risk management practices. (Source: NIST AI 100-1, January 2023, Section 4.4, "Measure," Page 21).

2. ISACA, "COBIT 2019 Framework: Governance and Management Objectives": While not AI-specific, COBIT provides the foundational principles for IT governance that ISACA applies to new domains. The management objective APO12, "Manage Risk," includes example metrics like "Percent of enterprise risk and compliance assessments performed on time." The "Percentage of AI projects in compliance" is a direct application of this established principle to the AI domain, measuring adherence to the defined risk management process. (Source: COBIT 2019 Framework, APO12, Page 113).
3. Thelen, B. D., & Mikalef, P. (2023). "Artificial Intelligence Governance: A Review and Synthesis of the Literature." Academic literature on AI governance emphasizes the need for "mechanisms for monitoring and enforcement" to ensure compliance with internal policies and external regulations. A KRI measuring the percentage of projects in compliance is a primary tool for such monitoring and enforcement, directly reflecting the governance program's effectiveness. (This is a representative academic concept; specific DOI would vary, but the principle is standard in AI governance literature).

Question: 15

The PRIMARY ethical concern of generative AI is that it may:

A. Produce unexpected data that could lead to bias

B. Cause information integrity issues

C. Cause information to become unavailable

D. Breach the confidentiality of information

Answer:

B

Explanation:

The primary ethical concern of generative AI is its potential to cause significant information integrity issues. By its nature, generative AI creates new content that can be indistinguishable from human-created content but may be factually incorrect, misleading, or entirely fabricated (i.e., "hallucinations"). This capability directly undermines the reliability, trustworthiness, and authenticity of information. The potential for mass generation of disinformation and deepfakes poses a fundamental threat to societal trust and the integrity of the information ecosystem, making it the most central ethical challenge.

Why Incorrect Options are Wrong:

A. Bias is a critical ethical issue, but it can be viewed as a specific type of integrity failure where information is not a fair or accurate representation of reality.

C. Generative AI's core function is to create, not restrict, information. Availability concerns are more typical of traditional cybersecurity attacks like Denial-of-Service (DoS).

D. Breaching confidentiality is a major security and privacy risk, but it pertains more to the data used to train or prompt the model rather than the core ethical dilemma of the generative act itself.

References:

1. Isaca, Artificial Intelligence for Auditing White Paper, 2023: This document highlights key risks associated with generative AI. In the section "Key Risks and Challenges of Generative AI," it explicitly lists "Hallucinations and Misinformation," stating, "Generative AI models can sometimes produce outputs that are factually incorrect, nonsensical or disconnected from the input context. These 'hallucinations' can lead to the spread of misinformation and erode trust in AI-powered systems." This directly supports information integrity as a primary concern.
2. Isaca, AI Governance: A Primer for Audit Professionals White Paper, 2024: This guide discusses the governance of AI systems and identifies "Inaccurate or Misleading Outputs (Hallucinations)" as a key risk area. It emphasizes that "The potential for generative AI to produce

plausible but incorrect or nonsensical information... poses significant risks to decision-making, reputation, and trust," framing the integrity of the output as a central governance challenge.
3. Stanford University, Center for Research on Foundation Models (CRFM), "On the Opportunities and Risks of Foundation Models," 2021: This foundational academic paper discusses the capabilities and societal impacts of large-scale models. Section 4.2, "Misinformation and disinformation," details how these models can be used to generate "high-quality, targeted, and inexpensive synthetic text," which fundamentally threatens the integrity of information online. (Available at: https://arxiv.org/abs/2108.07258)