# ISACA AAIA 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 AAIA Exam Questions by Cert Empire**

# Question: 1

Which of the following do supervised AI learning models PRIMARILY use to train algorithms?

A. Unlabeled data sets

B. Clustered data sets

C. Labeled data sets

D. Randomized data sets

## Answer:

C

## Explanation:

Supervised learning is a fundamental paradigm in machine learning where the algorithm learns from a dataset containing input-output pairs. The "label" is the correct output or target associated with each input data point. The primary goal of the model is to learn a general mapping function that can accurately predict the output for new, unseen inputs. This process is analogous to a student learning with a teacher who provides the correct answers (labels) for the problems (input data). The model's performance is evaluated based on its ability to correctly predict these labels.

## Why Incorrect Options are Wrong:



A. Unlabeled data sets: These are used in unsupervised learning, where the algorithm must find patterns or structures (like clusters) within the data without predefined correct answers.
B. Clustered data sets: Clustering is a common task within unsupervised learning. A clustered data set is typically the output of an algorithm, not the primary input for supervised training.
D. Randomized data sets: While data is often randomized or shuffled during the training process to prevent bias and improve generalization, this is a procedural step, not the defining characteristic of the data itself.

## References:

1. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press. In Chapter 1, Section 1.2, it states, "In the supervised-learning problem, the goal is to learn a mapping from inputs x to outputs y, given a labeled set of input-output pairs D = (xi, yi)Ni=1."
2. Ng, A. (2023). CS229 Machine Learning Course Notes. Stanford University. Part I, Supervised Learning, page 2, defines: "In supervised learning, we are given a data set and already know what our correct output should look like... we have a set of training examples, (x(i), y(i)) ; i = 1, . . . , m."
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. Chapter 2, Section 2.3, describes supervised learning as the process of learning from a training set of N example pairs (xi, yi), where yi is the label.

## Question: 2

From a data appropriateness and bias perspective, which of the following should be of GREATEST concern when reviewing an AI model used in a credit scoring system?

A. The model incorporates the applicant's loan history to assess spending habits.

B. The model utilizes historical credit data to predict future credit behavior.

C. The model considers the applicant's income level as a key factor in the credit decision.

D. The model uses postal codes as a primary factor in determining creditworthiness.

### Answer:

D

### Explanation:

Using postal codes as a primary factor is the greatest concern because it can serve as a strong proxy for protected characteristics such as race and ethnicity. This can lead to a form of systemic bias known as "digital redlining," where credit decisions are unfairly influenced by an applicant's geographic location rather than their individual financial merit. This practice can perpetuate and amplify historical patterns of discrimination, posing significant ethical and legal risks. The other factors, while needing careful handling, are direct, standard indicators of an individual's financial history and capacity to repay, which are legitimate considerations in credit scoring.



### Why Incorrect Options are Wrong:

A: An applicant's loan history is a direct and highly relevant measure of their past credit behavior, making it an appropriate and standard input for credit assessment.
B: Utilizing historical credit data is the fundamental principle of credit scoring models; it is the primary and most direct data for predicting future creditworthiness.
C: An applicant's income is a direct indicator of their financial capacity to repay a loan and is a legitimate and conventional factor in credit decisions.

### References:

1. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown. (While a book, its concepts are foundational in university courses on AI ethics). Chapter 1 discusses how zip codes are used as proxies for race, leading to discriminatory outcomes in models for insurance, credit, and parole.
2. National Institute of Standards and Technology (NIST). (2023). AI Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Section 4.3.2, "Systemic and Human Biases," discusses how proxy variables that correlate with protected classes can introduce harmful bias into AI systems.
3. Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. California Law Review,

104(3), 671-732. Page 677 discusses how seemingly neutral variables, such as zip codes, can function as proxies for protected classes like race, leading to discriminatory outcomes that may violate anti-discrimination laws.
4. Hardt, M. (2018). CS 294: Fairness in Machine Learning. University of California, Berkeley. Lecture 1 notes discuss the sources of bias and discrimination in machine learning, including the use of proxy variables like zip codes in credit scoring as a classic example of redlining. Available from UC Berkeley course materials.

# Question: 3

The PRIMARY objective of auditing AI systems is to:

A. Identify biases and decision transparency.

B. Maximize system efficiency and throughput.

C. Optimize user experience and interface satisfaction.

D. Minimize algorithm latency and information storage impacts.

## Answer:

A

## Explanation:

The primary objective of auditing AI systems is to provide assurance over their governance, risk management, and ethical operation. This fundamentally involves assessing whether the AI system functions fairly, without undue bias, and in a manner that is understandable to stakeholders. Identifying biases and ensuring decision transparency are core to verifying that an AI system aligns with organizational policies, ethical principles, and regulatory requirements. These elements are central to establishing trust and accountability, which are the ultimate goals of an AI audit, distinguishing it from performance testing or usability analysis.



## Why Incorrect Options are Wrong:

B: Maximizing system efficiency and throughput is a performance engineering objective, not the primary assurance-focused goal of an audit.
C: Optimizing user experience is a user-centered design (UCD) goal, which is separate from the audit's focus on risk, control, and compliance.
D: Minimizing latency and storage impacts are technical performance and resource management goals, not the primary risk and governance concerns of an AI audit.

## References:

1. ISACA. (2023). Auditing Artificial Intelligence. Page 11, Section "Key AI Audit Areas." The document explicitly lists "Fairness and Bias" and "Transparency and Explainability" as foundational areas for an AI audit, stating the audit should provide assurance that AI is used ethically and its risks are managed.
2. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., & Barnes, D. (2020). Closing the AI accountability gap: Auditing and public reporting as a route to accountability. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 332-344). The paper's central thesis is that AI auditing is a key mechanism for accountability, focusing on evaluating systems for bias, fairness, and transparency. https://doi.org/10.1145/3351095.3372873

3. The Alan Turing Institute. (2019). Understanding artificial intelligence ethics and safety. Page 12, Section "Principles for AI Ethics and Safety." This guide outlines core principles for trustworthy AI, including fairness and transparency, which are the primary subjects of an AI ethics audit.

## Question: 4

When auditing a machine learning (ML) solution, false positives can BEST be assessed by examining the level of:

A. Precision

B. Completeness

C. Accuracy

D. Recall

### Answer:

A

### Explanation:

Precision is the most direct metric for assessing the impact of false positives. It is calculated as the ratio of true positives to the sum of true positives and false positives (TP / (TP + FP)). A model that generates a high number of false positives will have a low precision score. Therefore, when an auditor's primary concern is to evaluate the rate of incorrect positive predictions (false alarms), examining the precision level is the most appropriate and specific method of assessment.



### Why Incorrect Options are Wrong:

B. Completeness: This is not a standard classification metric in machine learning. It is an ambiguous term in this context and likely a distractor.
C. Accuracy: Accuracy measures overall correctness ((TP+TN)/(TP+TN+FP+FN)) but can be misleading, especially in imbalanced datasets, as it does not specifically isolate the impact of false positives.
D. Recall: Recall (or sensitivity) measures the ability to find all actual positive instances (TP / (TP + FN)) and is primarily concerned with minimizing false negatives, not false positives.

### References:

1. Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies, 2(1), 37-63. (This foundational paper extensively defines and differentiates precision, recall, and other metrics, explicitly linking precision to the cost of false positives. See Section 2.1).
2. Stanford University. (n.d.). CS229 Machine Learning: Evaluation metrics. Course Notes. Retrieved from https://cs229.stanford.edu/notes2022fall/mainnotes.pdf (Section 4.3, "Precision and Recall," defines the metrics and their relationship to the confusion matrix, showing Precision's direct dependence on False Positives).
3. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8),

861-874. https://doi.org/10.1016/j.patrec.2005.10.010 (This paper, while focused on ROC, clearly defines the components of the confusion matrix, including False Positives, and how they relate to metrics like Precision and Recall on page 862).

## Question: 5

An IS auditor is performing an inventory audit for a manufacturing organization. Which of the following would BEST enable the auditor to identify types of products without assistance from organizational staff?

A. Natural language processing

B. Speech modeling

C. Robotic process automation (RPA)

D. Computer vision

### Answer:

D

### Explanation:

Computer vision is the field of artificial intelligence that enables computers to derive meaningful information from digital images, videos, and other visual inputs. In the context of an inventory audit, an auditor could use a computer vision system (e.g., an application on a mobile device) to automatically scan and identify different types of products based on their appearance, packaging, or barcodes. This technology directly addresses the need to classify physical objects visually, thereby allowing the auditor to perform the task without assistance from the organization's staff.



### Why Incorrect Options are Wrong:

A. Natural language processing: This technology processes and analyzes text and language, not visual data. It is unsuitable for identifying physical products from their appearance.
B. Speech modeling: This technology deals with analyzing and interpreting spoken language (audio). It has no application in a visual inventory audit.
C. Robotic process automation (RPA): RPA automates repetitive, rule-based digital tasks within software applications. It cannot interpret visual information to identify physical objects.

### References:

1. Stanford University. (n.d.). CS231n: Convolutional Neural Networks for Visual Recognition. Course Syllabus. "Computer vision is one of the fastest-growing and most exciting AI disciplines... This course is a deep dive into details of the deep learning architectures with a focus on learning end-to-end models for these tasks, particularly image classification." Retrieved from http://cs231n.stanford.edu/
2. Massachusetts Institute of Technology. (2022). 6.S191: Introduction to Deep Learning. MIT OpenCourseWare, Lecture 3: Computer Vision. The lecture defines computer vision and covers its core application in image classification, which is the fundamental task of identifying what an object is in an image.

3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539. This foundational paper describes how deep learning models, particularly convolutional neural networks (CNNs), have revolutionized computer vision tasks like object recognition (p. 438).

## Question: 6

An IS auditor notes that an AI model achieved significantly better results on training data than on test dat a. Which of the following problems with the model has the IS auditor identified?

A. Underfitting

B. Overfitting

C. Generalization

D. Bias

### Answer:

B

### Explanation:

Overfitting occurs when a machine learning model learns the training data too well, including its noise and random fluctuations, rather than the underlying generalizable patterns. This results in a model that achieves high accuracy on the training data but fails to perform well on new, unseen data (the test data). The significant discrepancy between the model's performance on the training set versus the test set is the classic indicator of overfitting. The model has essentially "memorized" the training examples instead of learning to generalize.



### Why Incorrect Options are Wrong:

A. Underfitting: An underfit model is too simple to capture the underlying structure of the data, resulting in poor performance on both the training and test data.
C. Generalization: This is the desired outcome, not a problem. It refers to a model's ability to adapt properly to new, unseen data, which is the opposite of what is described.
D. Bias: While related to model performance, bias refers to systematic errors or prejudice in a model's predictions. The specific problem of a performance gap between training and test data is overfitting.

### References:

1. Ng, A. (2023). CS229 Machine Learning Course Notes: Regularization and Model Selection. Stanford University. In the section on model selection, overfitting is defined as a model having low error on the training set but a very large error on the test set.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. Chapter 7, Section 7.1, "Introduction," discusses how a model can be overly complex and fit the training data too well, leading to poor prediction on new data, a phenomenon known as overfitting.
3. Guttag, J. V. (2016). 6.0002 Introduction to Computational Thinking and Data Science, Lecture 9. MIT OpenCourseWare. The lecture explains that overfitting happens when a model is fit too

closely to the peculiarities of a training set, resulting in poor performance on other sets of data.
4. Ying, X. (2019). An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series, 1168, 022022. Section 2.1 defines overfitting as a model that "performs well on the training data but does not generalize well to new data." (DOI: https://doi.org/10.1088/1742-6596/1168/2/022022)

## Question: 7

An AI social media platform uses an algorithm to increase user engagement that could unintentionally promote divisive content. Which of the following is the BEST course of action to mitigate this risk?

A. Introduce controls allowing individuals to customize content preferences.

B. Suspend the algorithm until concerns are addressed.

C. Obtain users' consent for the content they wish to view.

D. Regularly audit and adjust algorithms to reduce biases.

### Answer:

D

### Explanation:

The most effective and systemic approach to mitigate the risk of an AI algorithm unintentionally promoting divisive content is to address the algorithm's behavior directly. Regularly auditing the algorithm's outputs for unintended negative consequences (such as amplifying divisive content, a form of bias) and subsequently adjusting its parameters, training data, or objective function is a core practice in responsible AI governance. This creates a continuous improvement cycle that directly targets the root cause of the problem, rather than shifting responsibility to users or halting operations. This proactive and iterative process ensures the AI system's alignment with ethical guidelines and intended societal outcomes.



### Why Incorrect Options are Wrong:

A: This places the burden of mitigation on the individual user and does not correct the underlying systemic issue of the algorithm's behavior.
B: This is a reactive and drastic measure that stops the platform's core function, not a sustainable, long-term risk mitigation strategy.
C: User consent does not solve the algorithmic problem; users may not understand the downstream effects of their consent, and it fails to address the platform's responsibility.

### References:

1. National Institute of Standards and Technology (NIST). (2023). AI Risk Management Framework (AI RMF 1.0). The framework's MEASURE and MANAGE functions advocate for this approach. The MEASURE function involves "evaluating AI systems for... harmful bias" (p. 23, Section 4.3), and the MANAGE function involves deploying "a risk treatment, or mitigation, for documented risks" (p. 26, Section 4.4). This directly supports the cycle of auditing and adjusting.
2. Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. Communications of the ACM, 59(2), 56-62. https://doi.org/10.1145/2844110 This publication emphasizes the need for

"algorithmic accountability," which includes ongoing inspection and auditing of algorithms to understand their effects and correct for undesirable outcomes, such as the promotion of biased or harmful content.
3. MIT OpenCourseWare. (2020). 6.S092: Social and Ethical Responsibilities of Computing (SERC). Fall 2020. Lecture materials on "Fairness and Algorithmic Bias" discuss various methods for auditing algorithms to detect and mitigate biases, highlighting that technical adjustments to the model are a primary method for addressing such issues.

## Question: 8

Which of the following is the MOST important risk for an IS auditor to consider when reviewing the adoption of an AI system?

A. Costs associated with AI system maintenance

B. Immaturity of AI systems in the industry

C. Bias in AI system decision making

D. Resistance to the use of AI technology

**Answer:**

C

**Explanation:**

Bias in AI decision-making is the most critical risk for an IS auditor because it directly undermines the integrity, fairness, and reliability of the system's outputs. This can lead to significant negative consequences, including discriminatory outcomes, legal and regulatory non-compliance (e.g., with anti-discrimination laws), reputational damage, and flawed strategic decisions. Unlike other risks, bias can be deeply embedded, difficult to detect, and have widespread, systemic impact on individuals and the organization. From an audit and assurance perspective, ensuring the fairness and ethical operation of an AI system is a paramount governance objective.



**Why Incorrect Options are Wrong:**

A. Costs associated with AI system maintenance: This is a financial and operational risk, but it does not carry the same level of ethical, legal, and reputational severity as systemic bias.
B. Immaturity of AI systems in the industry: This is a valid concern and a root cause of many issues, but bias is a specific, high-impact consequence of this immaturity, making it the more precise and critical risk.
D. Resistance to the use of AI technology: This is a change management and organizational risk. While it can hinder adoption, it does not represent a fundamental flaw in the system's decision-making process like bias does.

**References:**

1. ISACA, Auditing Artificial Intelligence, 2021. In the section "AI Risks," bias is explicitly identified as a major risk category. The document states, "Biased results can lead to poor business decisions, financial loss, reputational damage, and/or legal and regulatory consequences." (Page 11). This highlights its significance over other operational or financial concerns.
2. ISACA, Artificial Intelligence: An Audit/Assurance Primer, 2023. This guide emphasizes the importance of auditing for fairness and bias. It notes that "Unintended bias in AI systems can lead to unfair or discriminatory outcomes, which can have serious ethical and legal implications."

(Section: "Key Risks and Challenges in AI"). This positions bias as a primary audit concern.
3. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91. This seminal academic paper demonstrates the severe real-world consequences of bias in AI systems, showing how leading commercial systems had significant accuracy disparities based on gender and skin type. This underscores the critical nature of bias as a risk. DOI: Not available, but accessible via http://proceedings.mlr.press/v81/buolamwini18a.html.
4. Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson. Chapter 28, "Philosophical Foundations," discusses the ethical implications of AI, including fairness and bias. The text explains that biased systems can perpetuate and amplify societal inequalities, making it a fundamental risk that transcends purely technical or financial considerations. (Chapter 28.2, "The Ethics and Risks of Developing Artificial Intelligence").

## Question: 9

An organization uses an AI image generation platform to create promotional materials. An IS auditor identifies that the platform includes copyrighted images in its training dat a. Which of the following is the auditor's BEST recommendation to address this issue?

A. Implement a manual review process to ensure no copyrighted images are used in generated outputs.

B. Use a platform that certifies the provenance and licensing of its training data.

C. Label all AI-generated images to disclaim the possibility of third-party content.

D. Suspend the use of the platform until the training data is sanitized.

### Answer:

B

### Explanation:

The most effective and proactive recommendation is to address the root cause of the risk: the use of improperly licensed training data. Selecting a platform that certifies the provenance and licensing of its data is a preventive control that ensures the organization complies with copyright law and adheres to ethical AI principles from the outset. This approach aligns with robust third-party risk management and AI governance by shifting the responsibility for data integrity to the vendor and providing a verifiable basis for compliance, which is a more sustainable and legally sound strategy than reactive measures.



### Why Incorrect Options are Wrong:

A: A manual review is a detective control that is inefficient, costly, and prone to human error. It does not address the fundamental risk of using a non-compliant platform.
C: Labeling outputs with a disclaimer does not absolve the organization of legal liability for copyright infringement and is not considered an effective risk mitigation control.
D: Suspending use is a temporary risk avoidance tactic, not a long-term solution. The best recommendation should guide the organization toward a sustainable, compliant operational model.

### References:

1. ISACA. (2023). Artificial Intelligence Audit Toolkit. The toolkit emphasizes the importance of data governance in AI systems. The "Data" domain section (Control Objective DA-02) specifically addresses the need to verify that "data acquisition and use are in compliance with legal, regulatory, and contractual requirements," which directly supports ensuring the provenance and licensing of training data.
2. National Institute of Standards and Technology (NIST). (2023). AI Risk Management

Framework (AI RMF 1.0). The "Govern" function of the framework highlights the importance of policies for data handling. Section 4.3.3, "Data," discusses the need to understand data sources, limitations, and provenance to manage AI risks effectively. Choosing a vendor that certifies this information aligns with this principle.
3. The Alan Turing Institute. (2019). Understanding artificial intelligence ethics and safety. This guide discusses the importance of data provenance as a cornerstone of responsible AI development. It notes that "the provenance of a dataset-its history and origins-is a key piece of information for assessing its quality and suitability for a given purpose" (p. 21), which includes legal and ethical suitability.

## Question: 10

An IS auditor uses an internally developed generative AI tool to prepare a status update for audit stakeholders. Which of the following is the auditor's MOST appropriate course of action?

A. Compare results with a publicly available generative AI tool to ensure outputs are similar.

B. Assess whether the information provided is complete and accurate.

C. Regenerate the results to ensure similar outputs are provided.

D. Share and review the results with management.

### Answer:

B

### Explanation:

The auditor's fundamental responsibility is to exercise professional due care, which includes ensuring that all communications and work products are accurate, complete, and supported by sufficient evidence. When using a generative AI tool, the output is merely a draft that assists the auditor; it is not a substitute for professional judgment and verification. The auditor remains fully accountable for the content they disseminate. Therefore, the most critical and appropriate action is to meticulously assess the AI-generated status update for factual accuracy and completeness against the underlying audit evidence before taking any further steps.



### Why Incorrect Options are Wrong:

A: Comparing with a public tool is inappropriate as it introduces data confidentiality risks and does not validate accuracy, since both tools could be flawed or biased differently.
C: Regenerating results only tests the model's output consistency (stability), not the factual correctness of the information provided in a specific instance.
D: Sharing results with management before the auditor has personally verified their accuracy and completeness is a failure of professional due care.

### References:

1. ISACA. (2023). Artificial Intelligence for Auditing. This white paper emphasizes the principle of human oversight, stating, "Auditors must understand the AI system's limitations and potential biases... and validate the outputs of AI systems to ensure their accuracy and reliability." This directly supports the need to assess the information before use.
2. ISACA. (2020). ITAF: A Professional Practices Framework for IS Audit/Assurance, 4th Edition. Standard 1204, Professional Due Care, requires IS auditors to exercise care and diligence. Guideline 2204 further specifies that auditors must obtain sufficient and appropriate evidence. An unverified AI output does not constitute sufficient or appropriate evidence; it must be validated by the auditor.

3. Moffitt, K. C., Rozario, A. M., & Vasarhelyi, M. A. (2018). Robotic Process Automation for Auditing. Journal of Emerging Technologies in Accounting, 15(1), 1-10. While focused on RPA, the principles extend to AI. The paper underscores that automation assists, but does not replace, the auditor's professional judgment and responsibility to validate the outputs of automated systems. (https://doi.org/10.2308/jeta-52047)

# Question: 11

Which of the following is the PRIMARY benefit of implementing a robust data governance framework specific to AI solutions in an organization?

A. It focuses on enhancing the accuracy and reliability of AI model predictions.

B. It accelerates AI implementation timelines by fully automating data preparation processes.

C. It fosters adherence to industry regulations while minimizing the risk of data breaches and privacy
violations.

D. It reduces the need for human oversight, ensuring seamless and autonomous data governance.

## Answer:

C

## Explanation:

The primary benefit of a robust data governance framework for AI is to establish the foundational policies, processes, and controls necessary for legal, ethical, and secure data management. This framework is crucial for ensuring compliance with regulations like GDPR and industry standards, thereby mitigating significant legal, financial, and reputational risks associated with data breaches and privacy violations. While improved model performance is a positive outcome, the core, overarching purpose of governance is to ensure responsible and compliant use of data, which is a prerequisite for any sustainable AI implementation.



## Why Incorrect Options are Wrong:

A: Enhancing model accuracy is a secondary benefit resulting from improved data quality, which is one component of a governance framework, not its primary, all-encompassing purpose.
B: A governance framework often introduces necessary checks and controls, which can initially slow down, rather than accelerate, implementation to ensure compliance and quality. It does not "fully automate" processes.
D: Data governance defines and structures human oversight through roles and responsibilities (e.g., data stewards, owners); it does not reduce or eliminate the need for it.

## References:

1. ISACA. (2023). Artificial Intelligence: An Audit and Assurance Framework. The framework emphasizes that AI governance is essential for managing risks, including "compliance with laws and regulations" and "privacy." It places governance as the umbrella under which other objectives, like performance, are managed. (See Chapter 2: AI Governance).
2. Tallon, P. P. (2020). Data Governance in the Age of AI. MIT Sloan Center for Information

Systems Research (CISR), Research Briefing, Vol. XX, No. 5. This briefing highlights that a key driver for data governance is "managing risk and ensuring compliance," which becomes even more critical with the scale and complexity of AI systems.
3. National Institute of Standards and Technology (NIST). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). The framework's "Govern" function is foundational and focuses on cultivating a risk management culture, which includes legal compliance and addressing negative impacts, directly aligning with option C. (See Section 4.1: The Govern Function).

## Question: 12

Which of the following BEST ensures that an AI system complies with user data ownership rights under privacy regulations?

A. Applying data clustering techniques to anonymize data sets

B. Enforcing strict data retention policies to limit storage duration

C. Implementing a transparent data consent management process

D. Regularly conducting AI system performance testing for accuracy

**Answer:**

C

**Explanation:**

A transparent data consent management process is the most fundamental and direct mechanism for ensuring compliance with user data ownership rights under privacy regulations like the GDPR. Such a process ensures that data subjects are clearly informed about how their data will be used by the AI system and are given the explicit ability to grant, manage, or revoke their consent. This directly addresses the core principles of lawfulness, fairness, and transparency, which are foundational to data ownership and control as mandated by major privacy frameworks.



**Why Incorrect Options are Wrong:**

A. Applying data clustering techniques to anonymize data sets: Anonymization is a data protection technique, but it does not address the user's fundamental right to consent to the collection and processing of their data in the first place.
B. Enforcing strict data retention policies to limit storage duration: While important for the "storage limitation" principle, this addresses only one aspect of data rights (deletion) and not the initial, ongoing consent for data use.
D. Regularly conducting AI system performance testing for accuracy: This relates to the AI model's functional quality and fairness, not the legal basis for processing user data or respecting their ownership rights.

**References:**

1. General Data Protection Regulation (GDPR), Article 7, "Conditions for consent": This article explicitly outlines the requirements for valid consent, stating it must be freely given, specific, informed, and unambiguous. It also mandates that the data subject shall have the right to withdraw their consent at any time, making a consent management process essential. (Official Journal of the European Union, L 119/1, 4.5.2016).
2. ISACA, "Auditing Artificial Intelligence," 2021: The guide emphasizes that a key risk in AI is the failure to comply with privacy regulations. It highlights the importance of controls that ensure a

proper legal basis for data processing, with consent being a primary example, and the need for transparency with data subjects. (Specifically, see sections on Data Privacy and Governance).
3. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". AI Magazine, 38(3), 50-57. https://doi.org/10.1609/aimag.v38i3.2741: This paper discusses the implications of GDPR for AI, underscoring the importance of transparency and lawful processing bases, such as consent, as central tenets for compliance when using personal data in algorithmic systems.

## Question: 13

When using off-the-shelf AI models, which of the following is the MOST appropriate way for organizations to approach vendor management?

A. Ensure a minimum of three quotes have been obtained for market research and comparison.

B. Establish responsibility and clear terms for model updates and support.

C. Only use models from vendors with globally recognized accreditation.

D. Use the vendor only if the contract has been reviewed by the information security department.

### Answer:

B

### Explanation:

Off-the-shelf AI models are dynamic and require continuous management throughout their lifecycle. Unlike traditional software, their performance can degrade over time due to "model drift" as real-world data changes. It is therefore most critical for an organization to contractually establish clear responsibilities and terms with the vendor for ongoing model updates, maintenance, and support. This ensures the model remains effective, secure, and compliant, addressing the unique operational risks associated with AI systems. This proactive governance is a cornerstone of managing third-party AI risk.



### Why Incorrect Options are Wrong:

A: Obtaining multiple quotes is a standard procurement practice for cost-effectiveness, not a specific strategy for managing the unique lifecycle risks of an AI model.
C: The field of globally recognized AI accreditation is still maturing; making this a strict requirement could be impractical and overly restrictive at present.
D: An information security review is a critical but standard due diligence step for any third-party software, not the most distinguishing or primary approach for AI vendor management.

### References:

1. NIST AI Risk Management Framework (AI RMF 1.0): The "Govern" function of the framework emphasizes establishing policies and procedures for third-party AI systems. Specifically, section 4.2.3, "Third Party Risk Management," highlights the need to "understand and manage the risks associated with third-party AI actors and entities across the AI lifecycle." This directly implies the need for clear terms on updates and support. (Source: NIST, AI RMF 1.0, January 2023, Page 21).
2. Academic Publication on AI Governance: In "A governance framework for the application of AI in an enterprise context," the authors discuss the importance of lifecycle management. They state, "The AI lifecycle does not end with deployment... Continuous monitoring of the model's

performance is necessary to detect model drift... and trigger retraining or replacement." This underscores the necessity of pre-defined vendor responsibilities for updates. (Source: Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Journal of Business Research, 98, 263-272. Section 4.3. DOI: https://doi.org/10.1016/j.jbusres.2019.01.032).

3. University Courseware on AI Risk: Materials on managing AI systems often differentiate them from traditional IT. The need for "continuous validation" and managing "technical debt" in machine learning systems is a key theme. This directly relates to the vendor's role in providing updates and support to prevent model degradation. (Source: Based on concepts taught in courses like Stanford's CS229: Machine Learning, which covers the practical lifecycle of ML models).

## Question: 14

Which of the following is the PRIMARY reason IS auditors must be aware that generative AI may return different investment recommendations from the same set of data?

A. Limitations can arise in the quantification of risk profiles.

B. Neural node access varies each time the process is executed.

C. Computational logic is based on probabilities.

D. Servers are reconfigured periodically.

**Answer:**

C

**Explanation:**

The core computational logic of most generative AI models is probabilistic, not deterministic. When given a prompt or a set of data, the model calculates a probability distribution over a vast range of possible next words or tokens. The final output is then generated by sampling from this distribution. Techniques like temperature scaling and nucleus sampling are used to control the randomness of this sampling process. Because the output is a result of probabilistic sampling, running the same query multiple times can produce different, yet plausible, results. An IS auditor must understand this inherent stochasticity to evaluate the model's consistency, reliability, and the associated risks for high-stakes applications like financial advice.



**Why Incorrect Options are Wrong:**

A. Limitations can arise in the quantification of risk profiles. This is a consequence of the model's variable output, not the fundamental reason for the variability itself.
B. Neural node access varies each time the process is executed. This is an inaccurate description of the inference process; the variability is primarily due to probabilistic sampling of the output, not random internal pathways.
D. Servers are reconfigured periodically. This is an operational infrastructure issue and is unrelated to the fundamental mathematical principles upon which the generative AI model operates.

**References:**

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. Chapter 3, "Probability and Information Theory," establishes the probabilistic framework that underpins modern machine learning models, including generative AI.
2. Stanford University. (2023). CS224N: Natural Language Processing with Deep Learning, Lecture 11: Language Models and RNNs Part 2. This lecture explains that language models work by producing a probability distribution over the vocabulary for the next word, P(xt x1, ..., xt-1), and

then sampling from this distribution to generate text. This sampling is the source of variability.

3. Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2019). The Curious Case of Neural Text Degeneration. Proceedings of the International Conference on Learning Representations (ICLR). This paper details sampling strategies (e.g., nucleus sampling) that are explicitly probabilistic and designed to control the randomness in text generation to produce higher-quality, non-deterministic outputs. (Available at: https://arxiv.org/abs/1904.09751)

## Question: 15

An organization shares an AI model with external partners. One partner reports that sensitive data has been inadvertently exposed through the model's outputs. Which of the following is the IS auditor's BEST recommendation?

A. Limit the model's outputs to anonymized results while investigating further.

B. Audit the data pipelines of all partners to identify the source of the leak.

C. Disable the shared model and notify partners of the potential breach.

D. Retrain the model immediately and implement privacy-preserving techniques.

### Answer:

C

### Explanation:

The highest priority in an active data breach scenario is immediate containment to prevent further unauthorized disclosure of sensitive information. Disabling the shared model is the most direct and effective containment action, immediately stopping the data leakage. Notifying partners is a critical and concurrent step in incident response, ensuring transparency, fulfilling potential contractual or legal obligations, and allowing partners to take their own protective measures. This approach aligns with established cybersecurity incident response frameworks, which prioritize containment before proceeding to eradication and recovery.



### Why Incorrect Options are Wrong:

A. Limit the model's outputs to anonymized results while investigating further.
This is an insufficient containment measure; anonymization can be flawed or reversed, and it does not address the root cause of the leak.
B. Audit the data pipelines of all partners to identify the source of the leak.
This is an investigative (detection and analysis) step that should occur after containment. It does not stop the ongoing data exposure.
D. Retrain the model immediately and implement privacy-preserving techniques.
This is a long-term remediation and recovery action. Retraining is not immediate and would allow the data breach to continue in the interim.

### References:

1. National Institute of Standards and Technology (NIST). (2012). Special Publication 800-61 Rev. 2, Computer Security Incident Handling Guide. Section 3.2, "Containment, Eradication, & Recovery," emphasizes that containment is a critical first step after detection to limit the scope and magnitude of an incident. Disconnecting the affected system (i.e., disabling the model) is a primary containment strategy.

2. Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2016). The Dangers of Machine Learning and a Plan to Mitigate Them. This paper discusses various attacks, including those that extract sensitive training data from a model's outputs. The severity of such data leakage underscores the need for immediate action to halt the model's operation once a leak is confirmed. The logical first response is to stop the system causing the harm. (Available via arXiv:1611.01646).
3. Carlini, N., et al. (2021). Extracting Training Data from Large Language Models. This research demonstrates the practical feasibility of extracting verbatim text sequences from a model's training data. The findings highlight that a model can act as a "lossy, compressed copy" of its training set, making immediate shutdown upon discovery of a leak the only prudent course of action to prevent further data extraction. (Proceedings of the 30th USENIX Security Symposium).