AIF-C01 AI Practitioner Free Sample (30 Questions)

Unofficial practice material. Not affiliated with or endorsed by Amazon.

This free sample contains 30 of the 865 questions from the full AIF-C01 AI Practitioner question bank.

Q1. A retailer wants software that adjusts product recommendations for each shopper based on browsing behavior, which changes daily. Why is a rule-based system poorly suited for this?

A) Writing rules for every possible behavior combination is impractical, while a machine learning model can learn preferences from behavioral data

B) A rule-based system can keep up if the team adds rules every week

C) Rule-based systems are easier to retrain when shopper behavior shifts

D) Shoppers will accept the same recommendations all year

Correct answer: A

Explanation: Behavior has too many combinations to capture in hand-written rules, and it changes constantly. An ML model learns preference patterns directly from behavioral data and can be retrained as behavior shifts. Why the others are wrong - B: Behavior combinations change daily and multiply across shoppers; adding rules weekly cannot keep pace with that scale of change.; C: Rule-based systems are not retrained; people must rewrite the rules by hand each time behavior shifts.; D: The scenario states behavior changes daily, so static recommendations would quickly become stale.

Q2. A bank wants a model that examines each transaction and decides 'fraud' or 'legitimate'. What type of machine learning problem is this?

A) Regression

B) Clustering

C) Binary classification

D) Reinforcement learning

Correct answer: C

Explanation: The model chooses between two predefined categories (fraud or legitimate), which is binary classification - binary because there are exactly two classes. Why the others are wrong - A: No continuous numeric value is being predicted.; B: Fraud and legitimate are predefined categories, not groups discovered from the data.; D: This is a single prediction per transaction, not sequential decision-making with rewards.

Q3. A team has training, validation, and test datasets. Which workflow uses them correctly?

A) Tune on the test set and report results on the validation set

B) Tune and compare models on the validation set, then evaluate the final model once on the test set

C) Train on the test set and tune on the training set

D) Use the same dataset for training, tuning, and final evaluation

Correct answer: B

Explanation: The correct workflow is: tune and compare models on the validation set, then evaluate the single final model once on the test set. Using the test set for tuning would leak information and make the final score optimistic. Why the others are wrong - A: Tuning on the test set leaks test information into model choices, so the reported score is no longer an unbiased estimate of real-world performance.; C: Training on the test set destroys its value as unseen evaluation data; the model must be trained on the training set.; D: Using one dataset for everything means the model is evaluated on data it already learned from, which hides overfitting.

Q4. A company wants AI to determine whether customer reviews are positive or negative. Which AI

application area does this belong to?

A) Natural language processing (NLP), which works with human language such as text

B) Computer vision, which works with images and video

C) Speech recognition, which converts audio to text

D) Reinforcement learning, which trains agents through rewards

Correct answer: A

Explanation: NLP is the AI application area concerned with human language: understanding and generating text and speech, including tasks like sentiment analysis, translation, and summarization. Why the others are wrong - B: Computer vision processes images and video, not the text of written reviews.; C: Speech recognition converts spoken audio to text; written reviews are already text.; D: Reinforcement learning is a training paradigm using rewards, not the application area for analyzing review text.

Q5. A company wants to search its photo archive for every photo in which one specific employee appears, starting from a single reference photo of that employee. Which AWS service provides this face search capability?

A) Amazon Rekognition

B) Amazon Textract

C) Amazon Transcribe

D) Amazon Lex

Correct answer: A

Explanation: Amazon Rekognition can index faces into a collection and then search other images for matching faces, which finds every photo in which the specified person appears. Why the others are wrong - B: Amazon Textract extracts text from documents; it cannot identify people in photos.; C: Amazon Transcribe converts speech to text and performs no image analysis.; D: Amazon Lex builds conversational chatbots and has no image recognition capability.

Q6. A company wants software that learns to optimize warehouse robot routes by trying different actions and receiving rewards for faster deliveries. Which type of learning fits?

A) Supervised learning

B) Unsupervised learning

C) Transfer learning

D) Reinforcement learning

Correct answer: D

Explanation: Learning by trying actions and improving based on rewards is reinforcement learning, the standard approach for optimizing sequential decisions like robot routes. Why the others are wrong - A: Supervised learning needs labeled examples of correct answers, not trial-and-error rewards.; B: Unsupervised learning finds structure in unlabeled data; it does not optimize actions through rewards.; C: Transfer learning reuses a model trained on one task for another; the scenario describes learning from rewards, not reusing a model.

Q7. A developer sends a very long document to a model and receives an error that the input exceeds the maximum token limit. What does this mean?

A) The input text is too long for a single request and must be shortened or split

B) The model has run out of training data

C) The model deleted the document from its training data

D) The model can only process documents it saw during training

Correct answer: A

Explanation: Every model has a maximum number of tokens it accepts per request. Exceeding it means the input must be reduced - summarized, truncated, or processed in chunks. Why the others are wrong - B: Training data is fixed long before deployment. A token-limit error concerns the current request's size, not training.; C: Token limits apply to the input of a single request, not to the model's stored training data.; D: Language models can process brand-new documents supplied as input; this error is about the document's length, not its familiarity.

Q8. A customer support team receives thousands of long chat transcripts each week and wants brief overviews of each conversation for quality review. Which use case fits best?

A) Forecasting next quarter's support ticket volumes

B) Producing word-for-word transcripts with no corrections needed

C) Generating full-length rewordings of each transcript

D) Generating concise summaries of long conversations

Correct answer: D

Explanation: Summarization is a core generative AI strength: condensing long documents or conversations into their key points, which fits the team's quality-review need. Why the others are wrong - A: Volume forecasting is predictive analytics; the team needs overviews of conversations, which is summarization.; B: Transcription is a different task, and transcripts always need review; the need here is brief overviews, not verbatim text.; C: Rewording at full length defeats the purpose; the team asked for brief overviews, which is exactly what summarization provides.

Q9. A team wants to use a generative model to classify support tickets into categories, and needs the classifications to be consistent. What should they understand first?

A) Setting the temperature to zero makes outputs fully deterministic

B) Running the same prompt twice is enough to confirm consistency

C) Model outputs can vary between runs, so the team should evaluate consistency on sample tickets before relying on it

D) Using the model only for drafting replies, not for classification

Correct answer: C

Explanation: Because generation involves sampling, outputs can differ across runs; responsible teams measure that variation on representative samples before deploying. Why the others are wrong - A: Zero temperature reduces variation but does not guarantee identical outputs; consistency must still be measured on samples.; B: Two runs cannot characterize variation; the team needs to evaluate outputs across representative sample tickets.; D: Classification is a legitimate use when consistency is evaluated; avoiding it sidesteps the requirement instead of addressing it.

Q10. A team wants a model to summarize product reviews and return the result as exactly three bullet points. Adding 'return exactly three bullet points' to the prompt is an example of:

A) A training example

B) A negative prompt

C) A system instruction to retrain the model

D) An output indicator specifying the desired format

Correct answer: D

Explanation: An output indicator tells the model how the answer should be shaped - here, the format is three bullet points. Why the others are wrong - A: A training example shows input-output pairs; this is a format instruction, not an example.; B: A negative prompt says what to avoid, not what format to use.; C: Nothing in the prompt retrains the model; prompts only guide a single response.

Q11. A company wants a generative AI assistant that answers questions about its internal policy documents. What must be ready FIRST?

A) Get approval of the project's budget for model API costs

B) The policy documents collected, digitized, and accessible to the application

C) Select the largest available foundation model for the assistant

D) Nothing; generative AI needs no data

Correct answer: B

Explanation: An assistant that answers from internal documents needs those documents gathered and accessible first; without the source material, it cannot give grounded answers. Why the others are wrong - A: Budget approval is a project formality; without the source documents the assistant cannot answer at all.; C: Even the largest model cannot answer from documents it cannot access.; D: Generative AI assistants need data to work from; claiming otherwise is false.

Q12. A company connects its support chatbot to the current product documentation so answers are based on official content instead of the model's memory. Which limitation of generative AI does this BEST address?

A) The model's tendency to answer in an inconsistent tone

B) Hallucinations and stale knowledge, by grounding answers in up-to-date company content

C) The high cost of inference for long documents

D) The cost of training the model from scratch

Correct answer: B

Explanation: Supplying current documents as context grounds the model's answers in real content, reducing invented facts and the problem of a fixed training cutoff. Why the others are wrong - A: Tone inconsistency is a prompting issue; connecting documents addresses factual accuracy, not stylistic consistency.; C: Adding documents increases token usage rather than reducing it; grounding addresses accuracy and stale knowledge, not cost.; D: Providing documents as context avoids training entirely, so training cost is not the limitation being addressed.

Q13. A company deploys the same support chatbot across its website and mobile app and wants its answers to stay consistent over time. Which practice BEST supports this?

A) Save successful prompts as reusable templates with fixed model settings, and review changes before updating them

B) Retrain the foundation model from scratch every week

C) Pin the model to a specific version so its behavior never changes

D) Retrain the model monthly on the latest support tickets so answers stay current

Correct answer: A

Explanation: Versioned, reusable prompt templates with fixed settings keep behavior stable across deployments, while review before changes prevents drift. Why the others are wrong - B: Weekly retraining is enormously expensive and would not stabilize day-to-day answers.; C: A pinned model version does not stop teams from changing prompts and settings; versioned templates are what keep behavior consistent.; D:

Retraining alters model behavior over time, which is drift; consistency comes from stable versioned prompts and fixed settings.

Q14. A company has two needs: quick first drafts of internal meeting notes, and polished summaries of customer contracts that lawyers will review. How should model choice differ between them?

A) Use the most expensive model for both, since internal notes deserve the best quality

B) Use a lower-cost model for the low-stakes internal drafts and a higher-quality model for the contract summaries

C) Use the higher-quality model for both, to keep the setup simple

D) Use the lower-cost model for both, accepting slightly weaker contract summaries

Correct answer: B

Explanation: Model choice should match the stakes: low-stakes drafts can use a cheaper, faster model, while high-stakes, lawyer-reviewed summaries justify a higher-quality model. Why the others are wrong - A: Using the most expensive model for low-stakes drafts wastes money with no meaningful benefit.; C: Using the expensive model for low-stakes drafts wastes money with no meaningful benefit.; D: Contract summaries are high-stakes and lawyer-reviewed; accepting weaker quality there risks errors caught too late.

Q15. A company has two kinds of knowledge: stable company values that rarely change, and daily inventory levels. Where should each live in a RAG-based assistant design?

A) Both should be baked into the model's weights through fine-tuning

B) Both should live in the knowledge base, since RAG can retrieve anything

C) Frequently changing facts like inventory belong in the retrievable knowledge base; stable concepts can live in prompts or the model

D) Inventory should be memorized by the model and values retrieved per query

Correct answer: C

Explanation: RAG's strength is fresh, changing data: inventory levels belong in the knowledge base where they can be updated. Stable values can be conveyed in prompts or model behavior without retrieval. Why the others are wrong - A: Baking daily-changing inventory into weights would require constant retraining.; B: Stable values do not need retrieval; putting everything in the knowledge base adds needless retrieval for content that never changes.; D: This reverses the logic: changing data should be retrieved, while stable guidance suits prompts or tuning.

Q16. A team uses Amazon Bedrock and wants to adapt a supported foundation model using its own labeled examples of ideal customer-service responses. Which Bedrock customization option should it use?

A) Continued pre-training with unlabeled data

B) Fine-tuning with the labeled examples

C) Changing the model's token limit

D) Continued pre-training on the support conversation transcripts

Correct answer: B

Explanation: Amazon Bedrock supports fine-tuning for certain models, where you provide labeled examples and AWS trains a customized copy of the model. Labeled ideal responses are exactly the input fine-tuning expects. Why the others are wrong - A: Continued pre-training on Bedrock uses unlabeled text for domain adaptation, not labeled examples of desired behavior.; C: The token limit is a model property, not a customization option, and changing it would not teach customer-service behavior.; D: Continued pre-training

on Bedrock uses unlabeled text for domain adaptation; it does not take labeled examples of desired behavior the way fine-tuning does.

Q17. A retailer must generate product descriptions for 500 products, each following the same structure: headline, three bullet features, and a call to action. Which technique is BEST?
A) A prompt template with variables for the product name and details
B) Fine-tuning a separate model for each product
C) Zero-shot prompting with no structure described
D) Random temperature sampling for variety
Correct answer: A
Explanation: A prompt template defines the headline-plus-bullets-plus-CTA structure once, with variables filled per product. This keeps all 500 descriptions consistent and the template easy to maintain. Why the others are wrong - B: Training 500 separate models is absurdly expensive; one template serves all products.; C: With no structure described, the model would produce inconsistent formats across the 500 products.; D: Random sampling increases variety, but the requirement here is consistency, not variety.

Q18. A company deploys an AI agent that books meeting rooms. The agent sometimes books the wrong room because it misreads the calendar. What is the MOST appropriate mitigation?
A) Remove all tool access so the agent cannot book anything
B) Delete the foundation model from the application
C) Add confirmation prompts and improve the agent's instructions and calendar-integration logic
D) Fine-tuning the foundation model on historical booking data
Correct answer: C
Explanation: Better instructions, a more reliable calendar integration, and a confirmation step before booking address the error while preserving the agent's useful capability. Why the others are wrong - A: Removing tool access turns the agent back into a chatbot and destroys the use case entirely.; B: The foundation model provides the agent's reasoning; the errors come from instructions and integration, not from using a model.; D: The errors come from instructions and calendar integration, not from the model's knowledge; retraining a foundation model would not fix misread calendar fields.

Q19. A company wants to evaluate the quality of its AI assistant's summaries of customer calls. Which evaluation approach captures nuance that automated scores often miss?
A) Human evaluation, where reviewers judge the summaries for accuracy and usefulness
B) Counting the number of words in each summary
C) Measuring the average length of the summaries
D) Counting the number of summaries produced per hour
Correct answer: A
Explanation: Human reviewers can judge faithfulness, tone, and usefulness of summaries, catching quality issues that automated metrics cannot detect. Why the others are wrong - B: Word count says nothing about whether a summary is accurate or useful.; C: Length says nothing about whether a summary is faithful or useful.; D: Throughput measures capacity, not the accuracy or usefulness of each summary.

Q20. A company offers a public AI email-writing assistant and must ensure it cannot be used to draft phishing emails. Which safeguard BEST enforces this?
A) A polite note in the system prompt asking users not to misuse the assistant
B) Using a smaller foundation model for the assistant

C) Guardrails with denied-topic and content filters that block phishing-related requests
D) Deleting each generated email after it is sent
Correct answer: C
Explanation: Preventing misuse needs enforcement, not requests: guardrails evaluate prompts and block categories of abuse such as phishing content, regardless of how the user phrases the request. Why the others are wrong - A: A polite instruction can be ignored or talked around; it is not a control.; B: A smaller model can still write a convincing phishing email; model size is not the safeguard.; D: Deleting the output after it is sent does nothing to prevent the misuse itself.

Q21. A team has deployed version 1 of its assistant and built a candidate version 2. It wants to know which performs better with real users before fully switching. Which method best answers this?
A) A/B testing: route a small percentage of real traffic to version 2, compare user-satisfaction and task metrics, then decide
B) Deploy version 2 to half of all users right away and keep the one that generates fewer support tickets
C) Deploy version 2 to everyone immediately without comparison
D) Compare the two versions' training loss only
Correct answer: A
Explanation: A/B testing measures real user outcomes on live traffic for both versions, giving the most reliable signal of which performs better in production. Why the others are wrong - B: Support-ticket counts are a noisy, lagging signal and the rollout is uncontrolled; proper A/B testing routes a small, measured fraction of traffic and compares satisfaction and task metrics before deciding.; C: Full deployment without testing risks a bad rollout.; D: Training loss does not reflect real user experience.

Q22. An insurer needs an underwriting assistant with a consistent, formal underwriting tone, and it must also quote today's rate tables, which change monthly. Which customization pairing fits?
A) Fine-tuning for the stable formal tone, plus RAG for the changing rate tables
B) RAG for the tone and fine-tuning for the rate tables
C) Fine-tuning for both the tone and the monthly rates
D) Prompt engineering for the rates and no customization for tone
Correct answer: A
Explanation: Fine-tuning shapes persistent behavior like tone, while RAG supplies current facts; swapping them would need constant retraining for the rates and fail to stabilize the tone. Why the others are wrong - B: This reverses the pairing: tone is stable behavior suited to fine-tuning, and rates are changing facts suited to retrieval.; C: Fine-tuning on monthly rates would require retraining every month as tables change.; D: Prompts alone cannot reliably hold a formal tone across all outputs, and they cannot supply changing rate data.

Q23. A team wants to measure potential bias in its training data and model predictions on AWS. Which service is designed for this?
A) Amazon Polly
B) AWS CloudTrail
C) Amazon SageMaker Clarify
D) Amazon Route 53
Correct answer: C
Explanation: SageMaker Clarify provides bias metrics for datasets and models, plus explainability, to help

teams detect unfairness. Why the others are wrong - A: Polly converts text to speech; it does not measure bias.; B: CloudTrail logs API activity for auditing; it does not compute bias metrics.; D: Route 53 is a DNS service with no role in bias measurement.

Q24. A team wants to check its lending dataset and its trained model for unfair bias. What does Amazon SageMaker Clarify measure?

A) Only the model's inference latency

B) Bias in both the training data and the trained model's predictions

C) The model's accuracy on a held-out test set

D) Whether the model complies with data-privacy regulations

Correct answer: B

Explanation: Clarify computes bias metrics at two stages: pre-training metrics that reveal imbalances in the dataset itself, and post-training metrics that reveal whether the model's predictions treat groups differently. Why the others are wrong - A: Latency is a performance concern; Clarify measures fairness, not speed.; C: Test-set accuracy measures predictive performance; Clarify's role here is measuring bias in the data and in the predictions, a separate concern.; D: Privacy compliance is a separate governance area; Clarify measures fairness and bias, not regulatory privacy posture.

Q25. A company creates an AI governance policy that lists which AI use cases are permitted, which are restricted, and which require executive approval before launch. What is the main purpose of this acceptable-use policy?

A) To set clear boundaries on how teams may use AI so that risky or unauthorized deployments are prevented

B) To speed up model training by removing data preprocessing steps

C) To specify the acceptable-use terms that third-party AI vendors must sign before procurement

D) To track each team's AI spending against its quarterly budget

Correct answer: A

Explanation: An acceptable-use policy is a governance control: it defines permitted, restricted, and prohibited AI uses and the approval path, keeping teams from deploying AI in risky or unapproved ways. Why the others are wrong - B: Data preprocessing choices do not depend on an acceptable-use policy; the policy governs usage, not training speed.; C: Vendor contracts are a procurement concern; this policy governs how internal teams may use AI.; D: Budget tracking is a cost-governance activity, not the purpose of an acceptable-use policy.

Q26. A company's AI governance policy requires annual reviews of all production AI systems for bias, safety, and compliance. Why is a recurring review cadence part of good governance?

A) Because models, data, and regulations change over time, so a one-time review cannot stay valid

B) Because a fixed review date satisfies auditors even if nothing has changed

C) Because review findings must be published publicly every year

D) Because each review replaces the previous one, so only the latest findings matter

Correct answer: A

Explanation: Governance is ongoing: data drifts, models degrade, and rules evolve, so periodic reviews are needed to keep AI systems fair, safe, and compliant over their lifetime. Why the others are wrong - B: Treating reviews as calendar theater misses the point: they matter because models, data, and regulations actually change.; C: Public reporting is a separate choice; the cadence exists because a one-time review

goes stale as conditions change.; D: Reviews build on history; dismissing earlier findings loses the trend the cadence is meant to track.

Q27. A company must move anonymized training data from the EU to the US for a training run. Which practice reduces the privacy risk of this transfer?

A) Transfer the full records over TLS but skip de-identification since the transfer itself is encrypted

B) Store the dataset in the US first, then de-identify it after the training run

C) Obtain customer consent forms before transferring the raw records

D) De-identify the dataset so no records are traceable to individuals before it crosses borders

Correct answer: D

Explanation: De-identification removes or masks the fields that could link records to individuals, so even if the transferred dataset is exposed, it cannot be traced back to specific people - the standard way to reduce privacy risk when data crosses borders. Why the others are wrong - A: TLS protects data in transit, but the dataset stays identifiable - anyone with access to the transferred data, or any later breach, still sees who each record belongs to.; B: De-identifying after the transfer leaves identifiable records exposed while they cross the border; it must happen before the data leaves the EU.; C: Consent addresses the legal basis for processing, but the transferred records remain identifiable - de-identification is what reduces the privacy risk of the transfer itself.

Q28. A company suspects a poisoning incident and needs to determine exactly which datasets trained each model version. Which capability helps answer this?

A) Prompt caching

B) Model quantization

C) Data lineage tracking, such as SageMaker ML Lineage Tracking

D) Increasing the inference timeout

Correct answer: C

Explanation: Data lineage records which data artifacts flowed into each training job, so investigators can trace a suspect model back to the exact dataset versions that produced it. Why the others are wrong - A: Prompt caching stores repeated responses for speed; it records nothing about training data.; B: Quantization shrinks model size and reveals nothing about data origins.; D: Timeout settings control request duration, not the history of training inputs.

Q29. A company must let a partner organization invoke its SageMaker real-time endpoint without giving the partner any access to its own AWS account. Which approach is most secure?

A) Issue long-lived IAM user access keys to the partner's developers for the endpoint

B) Have the partner assume a cross-account IAM role scoped to sagemaker:InvokeEndpoint, with the endpoint deployed in a VPC and encrypted in transit

C) Create IAM users for the partner's developers in the company's own account with the SageMakerFullAccess managed policy

D) Whitelist the partner's office IP range on the endpoint so it can be called directly

Correct answer: B

Explanation: A scoped cross-account role gives the partner only invocation rights, VPC placement removes internet exposure, and TLS encryption protects data in transit. Why the others are wrong - A: Long-lived shared credentials violate least privilege and cross-account boundaries; a scoped cross-account role is the secure pattern.; C: This grants the partner access inside the company's account with far broader permissions

than invoking one endpoint.; D: IP allowlisting is not authentication, and SageMaker endpoints require IAM authorization; the endpoint would still be exposed.

Q30. A company trains models on highly sensitive data and wants to ensure the training job cannot exfiltrate data over the internet. Which network design achieves this?

A) Allow internet access but block it with a security group during the training run

B) Run training in a VPC with no internet gateway or NAT, using VPC endpoints for S3 access only

C) Place the training job in a private subnet with a NAT gateway for package downloads

D) Mirror all training traffic to an inspection VPC for monitoring

Correct answer: B

Explanation: A VPC without internet egress removes the exfiltration path entirely, while S3 VPC endpoints preserve the only network access the job legitimately needs. Why the others are wrong - A: Security groups are reversible configuration, not a structural guarantee; a misconfiguration reopens the exfiltration path.; C: A NAT gateway provides internet egress, which is exactly the exfiltration path the design must eliminate.; D: Traffic inspection detects exfiltration after the fact; it does not prevent it.