Anthropic.CCAR-P.v2026-10-26.q56

Exam Code:	CCAR-P
Exam Name:	Claude Certified Architect - Professional
Certification Provider:	Anthropic
Free Question Number:	56
Version:	v2026-10-26
# of views:	102
# of Questions views:	561
https://www.freecram.net/torrent/Anthropic.CCAR-P.v2026-10-26.q56.html	

NEW QUESTION: 1
You are listing characteristics of strong architectural-decision communication.
Which two characteristics belong on the list? (Select two.)
Each correct answer presents a complete solution.
A. Ownership and review cadence for revisiting the decision when conditions or assumptions change.
B. Distribution to a standing review forum with mandatory attendance from senior engineering leaders.
C. Omission of rejected alternatives to keep the decision document focused and concise for readers.
D. Alternatives considered along with the criteria that were used to evaluate each alternative.
E. Use of a consistent template across the team to standardize the visual presentation of decisions.
Answer: (SHOW ANSWER)
Strong decision communication must explain both how the decision was reached and how it will remain governed. Option D provides the reasoning structure: the alternatives considered and the criteria used to compare them. This allows reviewers to verify that capability, cost, latency, security, maintainability, integration complexity, and business outcomes were evaluated consistently rather than selectively.
Option A establishes lifecycle accountability. Architectural decisions depend on assumptions about requirements, models, data, regulations, usage volume, and operational constraints. Naming an owner and review cadence ensures the decision is reconsidered when those assumptions materially change.
Option B imposes a meeting mechanism rather than describing decision quality. Mandatory attendance by senior leaders may add delay without improving the clarity or correctness of the record. Option C removes important evidence and increases the chance that rejected approaches will be proposed again without addressing the original objections. Option E can improve consistency and discoverability, but visual standardization alone does not establish substantive architectural communication.
A strong decision record should therefore be evidence-based, audience-accessible, traceable to requirements, explicit about alternatives and consequences, assigned to an accountable owner, and governed by defined review triggers.
Study Guide references/topics: Architectural-decision communication; alternatives analysis; evaluation criteria; decision ownership; review cadence; changing assumptions; lifecycle governance.

NEW QUESTION: 2
An engineering organization is adopting Claude Code across 200 developers. A team lead proposes that AI- generated pull requests bypass standard code review for changes under 50 lines because small changes are considered low risk and review capacity is constrained.
Which two Diligence-competency objections should you raise? (Select two.) Each correct answer presents part of the solution.
A. Removing review eliminates the verification mechanism required for AI-generated output across the workflow.
B. Code-review capacity should be expanded to handle every code change submitted across the team.
C. The threshold should be increased to 200 lines to capture more changes.

D. AI-generated pull requests should be rejected by default across the organization.
E. Line count is a poor proxy for the actual risk introduced by a code change.
Answer: (SHOW ANSWER)
AI-generated code requires verification through the same controlled engineering workflow as human-authored changes. Removing review eliminates the gate that can detect incorrect assumptions, security defects, missing tests, and unintended behavior. Line count is not a defensible risk measure: a one-line authorization change, dependency alteration, or command-execution defect can be more consequential than a large documentation update. Review requirements should instead be determined by affected components, privilege, data sensitivity, reversibility, test coverage, and deployment impact. Expanding capacity may be operationally useful but is not the central diligence objection. Raising the threshold retains the same flawed proxy, while rejecting all AI-generated changes discards legitimate productivity benefits. Anthropic states that users remain responsible for reviewing proposed code and commands. Claude Code security

NEW QUESTION: 3
You are identifying inefficiency in a Claude Code workflow where each engineer manually re-explains the project ' s conventions and architecture in every session.
Which adjustment most directly removes this inefficiency?
A. Forbid the use of Claude Code on the project entirely to avoid the session-initialization overhead, accepting that the team loses all AI-assisted development productivity for this codebase.
B. Capture the project ' s conventions and architecture in a project-scoped CLAUDE.md (or equivalent persistent project-context file) committed to the repository, so each session loads it automatically.
C. Tell each engineer to retype the project conventions and architecture context more quickly at the start of each session, reducing time lost without eliminating the repeated manual effort.
D. Remove all documented project conventions and architectural standards so engineers have nothing to re- explain at session start, accepting the loss of consistency and shared coding standards.
Answer: (SHOW ANSWER)
A project-scoped CLAUDE.md converts repeatedly supplied information into persistent, version-controlled context. Claude Code loads applicable CLAUDE.md instructions at session initialization, allowing architecture principles, coding conventions, build commands, repository structure, and verification requirements to be shared consistently across engineers. Anthropic explicitly recommends using CLAUDE.
md for information that developers would otherwise explain repeatedly; concise and specific instructions provide stronger adherence. Committing the file also enables normal review and change-control practices.
Option C reduces typing speed but preserves the underlying waste and inconsistency. Options A and D remove useful capability or governance rather than correcting the workflow. The implementation should periodically review the file for obsolete, conflicting, or unnecessarily verbose instructions.
Study Guide references/topics: Claude Code project memory; CLAUDE.md; persistent context; repository conventions; workflow standardization.

NEW QUESTION: 4
You are designing the prompt for a ticket-triage classifier with thirty well-defined categories and clear category descriptions in the prompt.
Which technique is most appropriate as the starting point?
A. A chain-of-thought prompt for a routing decision that does not require multi-step reasoning.
B. A zero-shot prompt that specifies the categories with their descriptions and the required output format.
C. A prompt that demands the model invent new categories when the supplied ones do not fit.
D. A prompt with no category descriptions, expecting the model to infer the category set.
Answer: (SHOW ANSWER)
A clearly defined closed-set classification task should begin with a concise zero-shot prompt containing the permitted categories, unambiguous category descriptions, and the required output structure. This provides a simple, inexpensive baseline that can be measured against a representative evaluation set. Chain-of-thought prompting adds unnecessary latency and token usage when the routing decision does not require substantive multi-step reasoning. Permitting Claude to invent categories would violate the closed-set contract and complicate downstream automation. Omitting descriptions would increase ambiguity among potentially overlapping labels. If evaluation later reveals persistent confusion between particular categories, the architect can introduce targeted examples or refine the definitions. Prompt complexity should be added in response to measured failure patterns, not assumed to be necessary before establishing baseline performance. Anthropic:
Prompt engineering overview

NEW QUESTION: 5
You are responding to a stakeholder request that adds significant scope to a deployment already in flight.

Which response best aligns expectations while respecting the stakeholder's underlying need?
A. Hide the scope-addition request from the broader delivery team to avoid a difficult conversation, allowing the change to proceed without team awareness of its timeline implications.
B. Decline the request outright without acknowledging the stakeholder's underlying need or offering alternatives such as deferral, descoping, or timeline extension.
C. Acknowledge the request, describe the scope and timeline implications transparently, and propose options-defer it, descope something else, or extend the timeline-for the stakeholder to choose.
D. Accept the additional scope into the current release without describing the timeline or descoping implications, allowing the delivery plan to absorb unacknowledged changes until they surface as delays.
Answer: (SHOW ANSWER)
Option C preserves both transparency and stakeholder agency. The architect acknowledges the underlying business need while making the delivery consequences explicit. Presenting controlled alternatives converts an unbounded request into a governed decision: defer the capability, exchange it for existing scope, or accept a later release date. The selected option should be documented with revised success criteria, dependencies, risk, cost, ownership, and approval. Concealing the request prevents accurate planning and creates unmanaged delivery risk. Rejecting it without exploring alternatives ignores legitimate stakeholder value. Quietly absorbing additional work produces scope creep and makes later delays appear unexpected. Effective lifecycle communication does not promise every requested capability; it presents the trade-offs necessary for an informed decision.
Study Guide references/topics: Scope governance; expectation alignment; transparent trade-offs; change control; stakeholder decision rights.

NEW QUESTION: 6
A Claude architect is designing a HIPAA-compliant pipeline that processes patient records.
Which two design decisions directly support HIPAA compliance requirements? (Select two.)
A. Setting max_tokens to a low value to minimize the volume of text generated per request.
B. Selecting the highest-capability Claude model to maximize diagnostic accuracy.
C. Enforcing role-based access controls so that PHI is retrievable only by authorized personnel.
D. Ensuring patient data is never included in training feedback loops sent to the model provider without a BAA in place.
E. Using streaming responses to reduce perceived latency for clinical users.
Answer: (SHOW ANSWER)
Role-based access control directly supports HIPAA's access-control and minimum-necessary principles. The retrieval and orchestration layers must verify the authenticated user's role before protected health information enters the model context. Authorization should be applied to individual records, data sources, tools, and actions, with audit events recording the requesting identity, accessed resource, purpose, and result. Model instructions alone are not a security boundary.
Patient information must also remain within services and processing activities covered by the organization's Business Associate Agreement. Anthropic states that PHI processing requires a HIPAA-ready organization and an accepted BAA, and that only specified eligible services and configurations are covered. Patient data must not be submitted through training, feedback, beta, third-party, or integration pathways that fall outside the applicable agreement. A BAA is not blanket authorization; architects must verify feature eligibility and disable inappropriate data-sharing mechanisms.
Reducing max_tokens, selecting a more capable model, and enabling streaming affect output length, quality, or latency. They do not establish authorization, contractual coverage, minimum-necessary access, or compliant PHI handling.
Study Guide references/topics: Anthropic Business Associate Agreement requirements; HIPAA-ready services; PHI access control; minimum necessary; eligible-service boundaries; auditability.

NEW QUESTION: 7
You are distinguishing functional from non-functional requirements during discovery.
Which item is a non-functional requirement?
A. The system must extract a defined set of specific fields from invoice attachments and populate a downstream data record.
B. The system must produce a draft response that a human reviewer can edit before sending.
C. The system must classify inbound tickets into a defined set of categories.
D. The system must respond at p95 latency under 800 milliseconds at the expected request volume.
Answer: (SHOW ANSWER)
Option D defines a measurable performance characteristic rather than a business capability. The requirement specifies a percentile-based latency threshold-under 800 milliseconds at the 95th percentile-and ties that threshold to the expected workload volume. It therefore describes how well the system must operate under defined conditions.
Options A, B, and C are functional requirements. They describe actions the solution must perform: extracting fields and creating a downstream record, producing an editable draft, and assigning tickets to categories.

Functional requirements define observable system behaviors or outputs.
A strong non-functional requirement must include a metric, threshold, measurement method, and operating condition. For Option D, the architect should further define whether latency is measured end-to-end or only at the model API, the test duration, concurrency, input-size distribution, cache state, timeout behavior, geographic region, and treatment of failed requests. Without those details, teams may report incompatible results while claiming to satisfy the same requirement.
Non-functional requirements for Claude solutions commonly address latency, availability, throughput, cost, privacy, security, recoverability, observability, scalability, and regulatory compliance.
Study Guide references/topics: Structured discovery; functional requirements; non-functional requirements; p95 latency; workload conditions; measurable acceptance criteria.

NEW QUESTION: 8
You are designing a human-in-the-loop validation workflow for a new Claude-based deployment and must complete the design steps before piloting the workflow.
Which two steps must be completed BEFORE piloting the workflow with a representative subset of traffic?
(Select two.)
Each correct answer presents part of the solution.
A. Define the sampling strategy and the escalation criteria at each oversight point in the pipeline.
B. Iterate the workflow design based on observed pilot findings before broader rollout to production.
C. Onboard the reviewer pool with role-specific training on the check criteria and escalation procedures.
D. Document the workflow with check criteria, escalation paths, and service-level agreements (SLAs) for each step.
E. Identify the decision points in the pipeline that require human oversight by impact and reversibility.
Answer: (SHOW ANSWER)
The workflow must first identify where human judgment is necessary. Option E applies risk-based oversight by examining each decision's impact and reversibility. High-impact, irreversible, externally binding, privacy- sensitive, or legally consequential actions require stronger review than low-risk draft generation or easily reversible recommendations.
After locating those control points, Option A defines how work reaches reviewers. The sampling strategy should specify whether review is universal, random, risk-triggered, confidence-based, or stratified. Escalation criteria should cover policy violations, uncertain outputs, conflicting evidence, sensitive data, failed validation, and cases exceeding the reviewer's authority.
Anthropic recommends human confirmation for decisions with meaningful real-world consequences and actions requiring affirmative consent. Computer Use Security Considerations Option B necessarily follows the pilot because findings must exist before design iteration. Reviewer onboarding and operational documentation in Options C and D are important readiness activities, but they follow the core design decisions being tested by this item: where oversight is required and how cases are sampled and escalated.
The completed pilot plan should subsequently include trained reviewers, documented procedures, queue capacity, SLAs, audit evidence, and measured inter-reviewer agreement.
Study Guide references/topics: Human-in-the-loop design; impact and reversibility; oversight points; sampling; escalation criteria; consequential decisions.

NEW QUESTION: 9
You are assessing a Claude-based system whose dominant risk is silent quality drift on safety-relevant outputs after a model-version upgrade.
Which assessment activity most directly addresses this risk?
A. Disable adversarial evaluation entirely during model-version upgrade cycles to reduce evaluation cost, accepting that safety drift will go undetected until it appears in production.
B. Rotate adversarial inputs randomly so no two upgrades are scored on the same set.
C. Skip evaluation on each model-version upgrade and rely on user-submitted complaints to surface safety drift after the upgraded model has already served production traffic.
D. Maintain an adversarial evaluation set with version-attributed scoring so each upgrade is measured against the same set before promotion.
Answer: (SHOW ANSWER)
A stable adversarial regression set provides a controlled basis for detecting behavioral change between model versions. Version-attributed results show whether the proposed upgrade improves, preserves, or degrades performance on the same safety-relevant cases. The evaluation must run before promotion so a regression can block deployment rather than becoming a production incident. Randomly replacing the test inputs removes comparability and makes score changes difficult to attribute. Disabling evaluation or relying on complaints converts a preventable release-control failure into reactive incident handling. Anthropic characterizes regression evaluations as tests of whether an agent still handles tasks it previously completed and recommends maintaining them continuously to detect backsliding. Demystifying evals for AI agents

NEW QUESTION: 10

The compliance team at a firm has approved a Claude Skill that generates client-facing investment summaries. The Skill includes the firm's required disclaimers and prohibited-language list. A product manager has asked whether additional guardrails are needed at the application layer or whether the Skill alone is sufficient.

Freecram.net

Which two guardrail responsibilities should remain at the application layer rather than the Skill? (Select two.) Each correct answer presents part of the solution.

A. Log every generated summary to the firm's compliance audit trail for retention.

B. Verify the requesting user is authorized to generate investment summaries at all.

C. Format output sections according to the firm's standardized house style guidelines.

D. Apply the disclaimer template that the compliance team has standardized firm-wide.

E. Apply the prohibited-language list that the compliance team maintains and updates.

Answer: (SHOW ANSWER)

Authorization and compliance logging must remain enforceable application-layer controls. Option B requires the application to authenticate the requester, resolve the user's role and entitlements, and deny the operation when the user lacks authority. A Skill can instruct Claude how to behave, but it is not a reliable security boundary and cannot replace identity-based access control enforced by the surrounding system.

Option A similarly belongs at the application layer. The application must create an authoritative audit event containing the requester, timestamp, Skill and model version, relevant input or content identifiers, generated output, review status, and retention classification. Audit evidence cannot depend on whether the model remembers or chooses to invoke a logging instruction.

Options C, D, and E describe reusable generation behavior that appropriately belongs within the approved Skill: house style, the standardized disclaimer, and the prohibited-language policy. The application may still validate their presence after generation, but the Skill remains the primary reusable content package.

The correct architecture separates behavioral guidance from deterministic enforcement. Skills supply domain instructions and templates; the application enforces identity, authorization, audit retention, data access, output validation, and consequential-action controls.

Study Guide references/topics: Skills versus application controls; authentication; authorization; compliance logging; deterministic guardrails; policy enforcement boundaries.

NEW QUESTION: 11

A Claude architect is implementing safety controls for a customer-facing advice assistant that must never provide regulated investment recommendations.

Which two guardrail implementations most directly enforce this constraint? (Select two.)

A. Increase response temperature to introduce variability that reduces the likelihood of specific recommendations.

B. Add an output classifier that detects and blocks responses containing regulated investment- recommendation language.

C. Limit session length to reduce the volume of queries processed per user per day.

D. Log all user queries to a SIEM for post-hoc compliance review.

E. Define explicit out-of-scope categories in the system prompt with fixed refusal phrasing for investment advice requests.

Answer: (SHOW ANSWER)

The required control is best implemented through layered prevention. The system prompt should explicitly classify personalized investment recommendations as out of scope, instruct Claude not to recommend particular securities or transactions, and provide standardized refusal language. Fixed refusal behavior reduces ambiguity and gives the model a clear alternative response when users request prohibited advice or attempt to reframe it indirectly.

An output classifier supplies an independent enforcement layer after generation but before delivery. It should detect prohibited recommendation language, individualized buy-or-sell instructions, portfolio allocations, and equivalent regulated content. A detected violation must block or replace the response rather than merely record it. This protects against prompt circumvention, unexpected phrasing, and occasional model noncompliance.

Increasing temperature makes responses less predictable and weakens control reliability. Session-length limits regulate volume, not content. SIEM logging is valuable for monitoring, investigation, and compliance evidence, but it is detective rather than preventative: the prohibited recommendation may already have reached the customer.

The completed design should evaluate both layers using adversarial prompts, paraphrases, multi-turn escalation, and classifier false-negative testing.

Study Guide references/topics: [Mitigating jailbreaks and prompt injections](https://docs.anthropic.com/en
/docs/test-and-evaluate/strengthen-guardrails/mitigate-jailbreaks); layered guardrails; explicit scope boundaries; standardized refusals; output classification; pre-delivery enforcement.

NEW QUESTION: 12
You are preparing a HIPAA-eligible deployment for a healthcare customer.
Which configuration supports HIPAA compliance using Anthropic-offered tools?
A. Claude Free with no contractual addendum, since consumer products meet HIPAA requirements out of the box.
B. Claude Enterprise with a signed Business Associate Agreement, Zero Data Retention enabled, and audit logging configured for compliance tracking.
C. Disabling all audit logging so that no PHI is recorded in any log store, on the assumption that the absence of logs satisfies HIPAA requirements without a signed BAA.
D. An ad-hoc personal Claude account used by individual clinicians for PHI-related tasks, with no Business Associate Agreement, no Zero Data Retention, and no audit logging configured.
Answer: (SHOW ANSWER)
Option B is the only configuration establishing the contractual, data-handling, and audit foundations needed for a HIPAA-eligible deployment. Anthropic requires an eligible commercial organization to activate the HIPAA-ready configuration and accept a Business Associate Agreement. A standard Enterprise subscription is not automatically covered merely because it is an enterprise product. Anthropic BAA Guidance Zero Data Retention may be required for particular covered services, including eligible Claude Code configurations, and is enabled per organization for qualified accounts. Audit logging supports access review, configuration monitoring, incident investigation, and compliance evidence. Anthropic provides audit logs for Enterprise organizations. Claude Code Legal and Compliance, Access Audit Logs Options A and D use consumer or personal accounts without the required contractual framework. Option C incorrectly treats the elimination of audit evidence as a substitute for a BAA and compliant configuration.
The signed BAA remains the authoritative source for covered products and features. The customer must also configure access control, minimum-necessary data use, workforce policies, retention, integrations, incident response, and vendor-risk controls. The configuration supports compliance; it does not independently guarantee it.
Study Guide references/topics: HIPAA readiness; BAA; ZDR; PHI handling; audit logging; eligible services; shared compliance responsibility.

NEW QUESTION: 13
A Claude-based research assistant begins producing responses that confidently contradict its retrieved source documents despite no change to the retrieval pipeline.
Which two diagnostic actions most directly identify the root cause of this behavior? (Select two.)
A. Increase the context-window size to allow more retrieved chunks per query.
B. Switch the retrieval index to a denser embedding model to improve chunk-relevance scores.
C. Determine whether the failure reproduces on the previous model version to test for a model mismatch.
D. Reduce the temperature setting to lower response variance across all query types.
E. Inspect the system-prompt grounding instructions to determine whether citation constraints remain intact.
Answer: (SHOW ANSWER)
The retrieval pipeline is stated to be unchanged, so the first investigation should isolate changes in generation behavior and grounding instructions. Replaying the failing cases on the previous model version determines whether the regression follows the model upgrade or current model-prompt combination. Inspecting the system prompt establishes whether source-use rules, uncertainty behavior, citation requirements, or prohibitions against unsupported claims were removed or weakened. Increasing context or replacing the embedding model changes a component for which no failure evidence exists and may introduce additional noise. Lowering temperature may reduce variation but does not explain systematic contradiction of retrieved evidence. The investigation should compare identical requests, retrieved passages, prompts, and model versions before applying a fix. Define success criteria and evaluations

NEW QUESTION: 14
You are reviewing a peer's end-to-end design for a Claude-based platform expected to scale to thousands of concurrent users.
For each statement, indicate Yes if it reflects sound architectural practice. Otherwise, select No.

Answer Area

Statements:	Select Yes or No:
Authentication and authorization run before retrieval so retrieval can filter by identity.	Yes / No
An asynchronous queue absorbs bursty traffic between intake and the model invocation layer.	Yes / No
Tax computation is encoded directly in the system prompt rather than in code.	Yes / No
Conversation logs include unredacted government identifiers to maximize signal for tuning.	Yes / No

freecram.net

Answer:

Answer Area

Statements:	Select Yes or No:
Authentication and authorization run before retrieval so retrieval can filter by identity.	Yes / No
An asynchronous queue absorbs bursty traffic between intake and the model invocation layer.	Yes / No
Tax computation is encoded directly in the system prompt rather than in code.	Yes / No
Conversation logs include unredacted government identifiers to maximize signal for tuning.	Yes / No

Explanation:
* Authentication and authorization run before retrieval so retrieval can filter by identity - Yes
* An asynchronous queue absorbs bursty traffic between intake and the model-invocation layer - Yes
* Tax computation is encoded directly in the system prompt rather than in code - No
* Conversation logs include unredacted government identifiers to maximize tuning signal - No Authentication and authorization must precede retrieval so unauthorized information never enters model context. An asynchronous queue is appropriate for burst absorption, backpressure, controlled concurrency, and retry management at scale. Deterministic tax computation should be implemented in validated code or a controlled calculation tool, not delegated to probabilistic prompt interpretation. Including unredacted government identifiers in conversation logs violates data-minimization principles and creates unnecessary privacy, security, and regulatory exposure. Sensitive fields should be removed, tokenized, or redacted before logging or model use unless specifically required and authorized. The resulting design separates deterministic computation, access control, scalable orchestration, and language reasoning into appropriate architectural layers.

NEW QUESTION: 15
You are auditing a procurement-assistant agent whose defined responsibility is to draft purchase requests for review.
For each tool currently configured on the agent, select yes if the tool should remain after a least-privilege audit. Otherwise, select no if it should be removed.

Answer Area

Tool:	Select Yes or No:
a read-only catalog lookup tool that returns vendor and SKU information	Yes / No
an inbox-send tool that emails purchase requests to vendors directly	Yes / No
a draft-write tool that creates purchase-request drafts in the review queue	Yes / No
a production-database admin tool unrelated to procurement	Yes / No
A read-only policy-document tool that returns the relevant procurement policies	Yes / No

Answer:

Answer Area

Tool:	Select Yes or No:
a read-only catalog lookup tool that returns vendor and SKU information	Yes / No
an inbox-send tool that emails purchase requests to vendors directly	Yes / No
a draft-write tool that creates purchase-request drafts in the review queue	Yes / No
a production-database admin tool unrelated to procurement	Yes / No
A read-only policy-document tool that returns the relevant procurement policies	Yes / No

freecram.net

Explanation:
Yes, No, Yes, No, Yes
Least privilege requires each retained tool to be necessary for the agent's assigned responsibility and restricted to the minimum action level needed. Catalog lookup is appropriate because vendor and SKU information is needed to prepare an accurate request. The policy-document tool is similarly justified because the agent must ground the draft in applicable procurement rules. The draft-write tool is also aligned because it writes only to a review queue, preserving the required human approval boundary.
The inbox-send tool exceeds the stated responsibility. Sending a request directly to a vendor converts a drafting assistant into an externally acting purchasing agent and bypasses the review stage. The production- database administrator tool is unrelated to procurement and introduces unnecessary destructive capability, privilege-escalation exposure, and potential data-access risk.
Anthropic's prompt-injection guidance recommends limiting access to sensitive data and actions and applying least privilege so that a compromised agent can cause minimal damage. Claude Code also supports tool- specific allow, ask, and deny rules that are enforced outside the model. Prompt-Injection Mitigation, Claude Code Permissions Study Guide references/topics: Capability-bloat auditing; least privilege; tool scoping; human review gates; separation of drafting and execution authority.

NEW QUESTION: 16
You are preparing an operational runbook for a Claude-based service.
Which content is essential to include in the runbook?
A. Dashboard and log references only, without alert definitions, triage steps, escalation paths, or rollback procedures for the on-call engineer to act on.
B. Common alerts and their triage steps, escalation paths, rollback procedures, and references to the relevant dashboards and logs.
C. Alert definitions and triage steps only, without escalation paths, rollback procedures, or references to dashboards and logs for on-call use.
D. Escalation paths and rollback procedures only, without alert definitions, triage steps, or dashboard references to guide initial incident response.

Answer: B (LEAVE A REPLY)
An actionable runbook must connect detection, diagnosis, containment, recovery, and escalation. Option B supplies that complete operational chain: alerts identify abnormal conditions; triage procedures guide initial investigation; dashboards and logs provide evidence; escalation paths assign authority; and rollback procedures restore a known-safe state. The runbook should cover model errors, latency degradation, rate limits, tool failures, prompt regressions, cost anomalies, data-governance events, and dependency outages.
Anthropic's operational materials demonstrate incident workflows in which alerts lead responders to logs and runbooks before corrective action. Options A, C, and D each omit information required for an on-call engineer to progress from notification to controlled resolution. Procedures should be tested through exercises and updated after incidents.
Study Guide references/topics: Claude operational and observability patterns; incident response; triage; escalation; rollback; operational readiness.

Valid CCAR-P Dumps shared by ExamDiscuss.com for Helping Passing CCAR-P Exam! ExamDiscuss.com now offer the **newest CCAR-P exam dumps**, the ExamDiscuss.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com CCAR-P dumps with Test Engine here: https://www.examdiscuss.com/Anthropic/exam/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 17
A Claude architect at a health services organization is defining evaluation metrics for a clinical-summary pipeline. The pipeline must remain within a per-query cost ceiling and must never surface patient data to unauthorized roles.
Which two metrics directly address these requirements? (Select two.)
A. BLEU score computed against a human-annotated reference summary set
B. Role-based access-control enforcement rate measured on a red-team dataset
C. Throughput measured as successful requests processed per minute
D. Per-query token cost measured against the defined cost ceiling
E. Response latency at the 95th percentile across a one-week sample window
Answer: (SHOW ANSWER)
The evaluation metrics must map directly to the two stated requirements. Role-based access-control enforcement rate, tested against a red-team dataset, measures whether requests from unauthorized identities, roles, or access paths are consistently denied. It converts the privacy and authorization requirement into an observable security metric and exposes bypass conditions that ordinary functional tests may miss. Because the requirement says patient data must never reach unauthorized roles, the acceptance threshold should be effectively 100% enforcement across representative and adversarial cases.
Per-query token cost directly measures compliance with the financial constraint. The metric should include uncached input, cache reads and writes where applicable, output tokens, thinking tokens, and any other billable components attributed to the request. Results can then be compared with the defined ceiling across normal, boundary, and worst-case inputs.
Anthropic's evaluation guidance identifies privacy preservation and price as distinct success criteria and emphasizes that criteria must be specific, measurable, and relevant to the application. Define Success Criteria and Build Evaluations BLEU may assess textual similarity but does not test authorization or cost. Throughput and percentile latency are valuable operational metrics, yet neither directly answers the two requirements in the scenario.
Study Guide references/topics: Multidimensional evaluation; privacy preservation; authorization testing; red- team datasets; token-cost measurement; requirement-to-metric traceability.

NEW QUESTION: 18
You are supporting a team whose Claude Code sessions consistently load 60 or more MCP tools from many servers, exhausting context budget before the session begins.
Which adjustment most directly addresses this issue without removing capability?
A. Enable Tool Search so tool definitions are deferred and discovered on demand rather than loaded into context upfront.
B. Disable every MCP server in the configuration to free up context budget, accepting that the team loses all tool access and cannot perform any MCP-dependent task in the session.
C. Increase the prompt's verbosity with additional instructions and context, which consumes more of the context budget rather than reducing the tool-definition overhead causing the issue.
D. Add additional MCP servers to give the team more capability, which increases rather than reduces the number of tool definitions loaded into the session's context budget.
Answer: (SHOW ANSWER)

Tool Search applies progressive discovery to a large tool catalog. Instead of inserting every MCP tool's complete JSON schema and description into the initial context, Claude receives limited discovery information and loads the full definitions only when a task requires them. This preserves the available context for instructions, repository content, conversation history, and tool results without removing the underlying capabilities.
Anthropic's Claude Code documentation states that Tool Search keeps MCP context usage low by deferring tool definitions until Claude needs them. Only the tools actually selected enter the working context. Claude Code MCP Tool Search Anthropic's advanced-tool-use guidance also describes conventional multi-server configurations consuming tens of thousands of tokens before work begins, while on-demand discovery substantially reduces that overhead.
Advanced Tool Use Option B solves context pressure by destroying required functionality. Option C worsens the constraint by adding more prompt tokens. Option D expands the exact source of context bloat.
The team should also improve server and tool descriptions so discovery returns the correct tools, monitor tool- selection quality, and keep frequently required foundational tools non-deferred only when evaluations justify the additional context cost.
Study Guide references/topics: MCP scaling; Tool Search; progressive discovery; deferred schemas; context- budget management; tool-selection accuracy.

NEW QUESTION: 19
You are reviewing a peer's Claude Code permission rules for an enterprise rollout. The rules grant unrestricted Bash access to all projects across all developers.
Which response is most appropriate?
A. Add unrestricted access to additional tool categories as well, so that Bash is not asymmetrically more permissive than other tools, expanding the attack surface further in the name of consistency.
B. Approve the unrestricted Bash access as written on the grounds that narrowing the rules would add configuration complexity, accepting the full attack surface for all engineers across all projects.
C. Replace unrestricted Bash with narrowly scoped tool patterns that allow only the specific commands the workflows require, and add explicit deny rules for sensitive operations.
D. Disable all permission rules for the enterprise rollout so every command across every project runs without any tool-pattern scoping or explicit deny rules for sensitive operations.
Answer: (SHOW ANSWER)
Unrestricted Bash access exposes filesystem, credential, process, network, package-management, and potentially infrastructure-administration capabilities across every repository. Option C applies least privilege by identifying the commands genuinely required by approved workflows and permitting only those patterns.
Explicit deny rules should protect destructive commands, sensitive files, credential stores, deployment systems, and other prohibited operations.
Anthropic's permission system supports allow, ask, and deny rules, with deny evaluated before ask and allow.
It also warns that command patterns intended to constrain network utilities can be fragile, so stronger controls such as WebFetch domain restrictions, hooks, and sandbox boundaries may be required. Claude Code Permissions Enterprise configuration should combine managed baseline restrictions with project-specific permissions where necessary. High-impact commands should remain subject to confirmation, and sandboxing should provide operating-system-level filesystem and network containment.
Options A and D deliberately enlarge the attack surface. Option B treats configuration simplicity as more important than enterprise security and fails to account for prompt injection, accidental commands, compromised repositories, or excessive human approval. The configuration should be narrow, auditable, centrally governed, and validated against real development workflows.
Study Guide references/topics: Claude Code permissions; Bash restrictions; least privilege; managed controls; explicit deny rules; sandboxing; enterprise rollout security.

NEW QUESTION: 20
You are diagnosing a Claude Code session whose subagent uses 50,000 tokens of context before the engineer types a single message.
Which root cause is most likely?
Freecram.net
A. The developer ' s keyboard layout or input-method configuration is the cause of the elevated context consumption, introducing extra tokens before the engineer types any message.
B. Many MCP servers are configured, each contributing tool definitions to the context budget; Tool Search is not enabled, so all definitions load upfront.
C. The model has internal personal preferences or default behaviors that silently consume large portions of context budget before any user message is processed, independent of tool configuration.
D. The font rendering or display-scaling settings of the IDE are converting visual output into additional context tokens, causing the high pre-session context consumption.
Answer: (SHOW ANSWER)
Claude Code constructs the model context before the first user message. That initial context can include the system prompt, project instructions, subagent configuration, MCP server instructions, and tool definitions expressed as JSON schemas. When numerous MCP servers expose many tools and Tool Search is disabled, those schemas must be loaded upfront. A large collection of verbose tool descriptions and parameter schemas can therefore consume tens of thousands of tokens before the engineer enters any text.

Tool Search is specifically designed to prevent this form of context inflation. With Tool Search enabled, Claude initially receives only server instructions and compact tool-identification information; full tool schemas are deferred until a relevant tool is discovered and required. Anthropic consequently describes Tool Search as the mechanism for keeping MCP context consumption low as integrations scale.
Keyboard layouts, font rendering, and display scaling operate outside the language-model context and cannot generate API input tokens. Likewise, model behavior does not silently reserve 50,000 tokens for personal preferences. The diagnostic action is to inspect configured MCP servers and tool counts, determine whether Tool Search has been disabled, and remove or defer unnecessary definitions.
Study Guide references/topics: [MCP Tool Search and context scaling](https://docs.anthropic.com/en/docs
/claude-code/mcp); MCP server configuration; subagent context budgeting; tool-schema optimization.

NEW QUESTION: 21
An architect at Trenova Systems, Inc. is reviewing two draft communication packages for the same architectural decision. Package 1 contains a threat model, residual-risk register, and control-mapping table.
Package 2 contains a list of capabilities delivered and feature-level roadmap dependencies.
Which two audiences are correctly matched to these packages? (Select two.)
A. Package 1 # engineering implementation team
B. Package 2 # executive sponsor
C. Package 1 # security and compliance reviewer
D. Package 2 # product manager owning the roadmap
E. Package 2 # security and compliance reviewer
Answer: (SHOW ANSWER)
Package 1 is designed for security and compliance review because its threat model identifies relevant threats and attack paths, its residual-risk register documents the risks remaining after controls, and its control- mapping table connects safeguards to organizational or regulatory requirements. Package 2 is designed for the product manager because delivered capabilities and roadmap dependencies directly support prioritization, release planning, sequencing, and lifecycle ownership. An engineering implementation team would ordinarily need component diagrams, interface contracts, configuration schemas, deployment instructions, and operational runbooks rather than primarily risk-governance artifacts. An executive sponsor generally needs a concise outcome, cost, risk, and decision summary instead of feature-level dependency details. Effective architecture communication must therefore adapt the content and level of detail to each audience's responsibilities and decision rights.

NEW QUESTION: 22
You are supporting an engineer whose Claude Code session reports a permission denial when the engineer expected the action to be allowed.
Which resolution step is most appropriate?
A. Tell the engineer to retry the denied action verbatim repeatedly until the denial stops appearing, without inspecting the active permission rules to determine whether the denial is intentional.
B. Disable the permission system entirely across all scopes to remove the denial, eliminating all tool- pattern and deny-rule controls rather than identifying and amending the specific rule.
C. Inspect the active permission rules across all scopes-managed, command-line, local, project, and user-to identify which rule is denying the action and confirm whether it should be amended or remain denied.
D. Grant the engineer unrestricted Bash access to bypass the specific rule causing the denial, removing all tool-pattern constraints rather than amending only the rule in question.
Answer: (SHOW ANSWER)
Claude Code permission decisions can be affected by multiple configuration scopes. The appropriate diagnostic action is to inspect the effective rules and identify the exact deny pattern. Managed policy has the highest precedence, and a deny rule from any applicable scope cannot be overridden by an allow rule elsewhere. The architect must then determine whether the denial represents an intentional security boundary or an incorrectly scoped rule. If correction is justified, only the specific rule should be amended. Repeated retries cannot change a deterministic permission decision. Disabling permissions or granting unrestricted Bash access removes safeguards far beyond the reported problem and violates least privilege. Anthropic's permission documentation explicitly describes scope precedence and deny-first evaluation. Claude Code permissions

NEW QUESTION: 23
You are transitioning a Claude-based deployment from design into implementation.
Which handoff package most directly supports a clean transition?
A. The most recent set of design presentation slides without component-level diagrams, interface contracts, an evaluation framework with a reference set, runbooks, or a known-limitations register.
B. A verbal walkthrough conducted on the day of handoff with no written architecture overview, ADRs, component contracts, evaluation framework, runbooks, playbook, or known limitations.

C. Architecture overview, ADRs, component contracts, evaluation framework with reference set, runbooks, on-call playbook, and known limitations.
D. Source code alone with no integrating architecture overview, ADRs, component contracts, evaluation framework, runbooks, on-call playbook, or known-limitations register to support the delivery team.
Answer: (SHOW ANSWER)
Option C provides the minimum coherent package required to move from architectural intent to implementable and supportable delivery. The architecture overview establishes system boundaries and component relationships.
Architecture Decision Records preserve the context and trade-offs behind key choices. Component contracts define interfaces, schemas, responsibilities, failure behavior, and integration assumptions.
The evaluation framework and reference set allow the implementation team to confirm that prompts, models, retrieval components, and tool integrations continue to meet the defined quality, safety, latency, and cost thresholds. Runbooks translate operational knowledge into repeatable procedures, while the on-call playbook defines detection, escalation, containment, and recovery actions. The known-limitations register prevents unsupported behavior from being mistaken for an implementation defect and helps stakeholders manage residual risk.
Options A and B depend excessively on slides or undocumented verbal knowledge. Option D provides implementation material without explaining the system-level decisions, operating model, acceptance criteria, or limitations. Source code cannot substitute for architectural and operational documentation.
A clean transition does not mean the architect can never be consulted. It means routine implementation and support do not depend on undocumented knowledge held by a single individual.
Study Guide references/topics: Design-to-implementation handoff; ADRs; component contracts; reference evaluations; runbooks; operational playbooks; limitations management.

NEW QUESTION: 24
You are reviewing a customer-support agent's configuration. Each candidate tool falls into one of four categories: (1) required to complete defined tasks, (2) frequently used and reduces hand-offs, (3) occasionally useful for unrelated work, (4) speculative future utility.
Which categories should typically remain in the agent configuration?
A. Only category 3, because occasionally useful tools for unrelated work provide broader coverage and should take priority over tools required for the agent's defined tasks.
B. Only category 4, because speculative future-utility tools provide the most flexibility and should be configured even when no defined task currently requires them.
C. Categories 1 and 2 only, because they map to defined tasks and the agent's regular hand-offs.
D. All four categories, because broader tool access is categorically better for agent performance regardless of whether the tools map to defined tasks or regular hand-offs.
Answer: (SHOW ANSWER)
Category 1 tools are necessary for the agent to complete its approved responsibilities and must remain.
Category 2 tools are also justified when they are regularly used and eliminate predictable hand-offs without expanding the agent beyond its defined operating role.
Categories 3 and 4 create capability bloat. A tool that is occasionally useful only for unrelated work does not belong to this agent's responsibility boundary. A speculative tool has no validated requirement and adds attack surface, context consumption, authorization complexity, evaluation burden, and operational dependencies without demonstrated value.
Tool inclusion should be based on task traceability: every configured capability should map to a documented user journey, responsibility, permission scope, and evaluation case. Anthropic recommends least privilege so that a successful injection or model error can cause minimal damage. Mitigate Jailbreaks and Prompt Injections The architect should periodically review tool-call telemetry and remove unused capabilities. New tools can be added when a validated workflow requires them, after security review and evaluation. The objective is not the smallest possible catalog regardless of usefulness; it is the smallest catalog that reliably completes the agent's approved tasks.
Study Guide references/topics: Capability-bloat analysis; tool necessity; hand-off reduction; least privilege; attack-surface management; tool lifecycle review.

NEW QUESTION: 25
When communicating an architectural decision to a security and compliance reviewer, which content set is most aligned with that audience's primary concerns?
A. Feature delivery schedule, roadmap dependencies, and scope boundaries for the product manager.
B. Implementation timeline, component interfaces, and deployment sequence for the engineering team.
C. Threat model, control mappings, residual-risk acceptance criteria, and audit traceability.
D. Decision rationale, business outcomes, and high-level risk summary for executive review.
Answer: (SHOW ANSWER)
Security and compliance reviewers must determine whether material threats have been identified, whether required controls address those threats, how remaining exposure will be accepted, and whether the decision can be demonstrated during an audit. Option C supplies exactly that evidence.

The threat model establishes protected assets, trust boundaries, adversaries, attack paths, and likely failure modes. Control mappings connect technical and procedural safeguards to organizational policies, regulatory requirements, and assurance frameworks. Residual-risk acceptance criteria define which risks may remain after mitigation, who has authority to accept them, and when further treatment is mandatory. Audit traceability connects requirements and decisions to implementation evidence, evaluations, approvals, logs, and operational monitoring.
Option A is designed primarily for product management. Option B addresses engineering planning and execution. Option D is appropriate for an executive sponsor who requires business justification and a concise risk summary but normally does not perform detailed control verification.
Audience adaptation should not alter the architectural decision itself. It changes the depth, terminology, evidence, and ordering used to communicate the decision. Security and compliance reviewers require verifiable control evidence rather than only schedule, implementation, or business-level summaries.
Study Guide references/topics: Audience-specific communication; threat modeling; compliance control mapping; residual-risk acceptance; audit evidence; architectural decision traceability.

NEW QUESTION: 26
You are defining where human review must remain in a planned automated pipeline. Which placement reflects sound human-in-the-loop design?
A. Place a human reviewer only after the irreversible action has already executed, making the review a post-hoc audit rather than a meaningful pre-action check or approval gate.
B. Place a human reviewer between the model's output and any high-impact, irreversible action, with explicit criteria for what the reviewer must check before approval.
C. Place a human reviewer in the loop for a randomly selected sample of requests, without defining criteria for what the reviewer should check or which output categories require mandatory review.
D. Remove all human review steps from the pipeline to maximize throughput, accepting that high-impact and irreversible actions will be taken without any human approval or oversight.
Answer: (SHOW ANSWER)
Sound human-in-the-loop design places review before a high-impact or irreversible action occurs. The reviewer must receive sufficient context and explicit approval criteria so that the checkpoint functions as a meaningful control rather than a ceremonial step. This is especially important when Claude outputs could initiate financial transactions, publish sensitive material, alter access rights, affect employment decisions, or trigger other consequential actions. Option A provides only retrospective auditing and cannot prevent harm.
Option C may support monitoring or quality measurement, but random sampling does not protect every case that requires mandatory approval, particularly when review criteria are undefined. Option D removes the accountability and intervention mechanism entirely. Therefore, B provides the strongest design by combining risk-based placement, human authority, and clearly defined verification requirements before execution.

NEW QUESTION: 27
You are assessing data-exfiltration risk in a Claude-based assistant that has tools for both internal-document retrieval and outbound HTTP calls.
Which scenario most directly indicates a data-exfiltration risk?
A. The outbound HTTP tool returns a 200 response when the assistant calls an allow-listed domain to fulfil a user-initiated data-lookup request.
B. The user submits a routine question and receives a routine answer.
C. Adversarial content in a retrieved document instructs the model to call the outbound HTTP tool and send sensitive content to an attacker-controlled URL.
D. The internal-document retrieval tool returns a document that the user is authorized to view.
Answer: (SHOW ANSWER)
Option C describes indirect prompt injection combined with an outbound transmission channel. The attacker embeds instructions in content that the assistant treats as evidence. If the model follows those instructions, it can retrieve sensitive information and transmit it through the HTTP tool to an attacker-controlled destination.
Authorization to read a document does not automatically authorize external disclosure. The architecture should treat retrieved content as untrusted, restrict outbound domains, separate data access from external communication, apply least-privilege tool permissions, and require approval for sensitive transmissions.
Options A, B, and D describe normal authorized behavior without evidence of unintended disclosure.
Anthropic specifically warns that combining untrusted content, sensitive data, and network tools creates exfiltration risk. Web-fetch security guidance; handling untrusted tool results

NEW QUESTION: 28
You are producing an architecture guide for a new deployment and must complete the planning steps before drafting each section.
Which two steps must be completed BEFORE drafting each section of the guide? (Select two.) Each correct answer presents part of the solution.
A. Identify the audience and the questions the guide must answer for that audience.

B. Translate the guide into the supported regional languages for the candidate population.
C. Validate the guide with the implementation team and incorporate corrections.
D. Establish the document under version control with a defined review cadence and approver list.
E. Outline the guide sections covering the overview, components, contracts, flows, runbooks, and limitations.
Answer: (SHOW ANSWER)
Effective documentation begins with audience analysis and structural planning. Identifying the audience determines the required technical depth, terminology, decision context, and operational emphasis. Defining the questions the guide must answer prevents sections from becoming generic descriptions. The outline then establishes complete coverage across architecture overview, component responsibilities, interface contracts, dominant flows, operational runbooks, and known limitations. Translation occurs only after stable source content exists. Implementation-team validation also follows an initial draft because there must be material to review. Version control, reviewers, and cadence are important governance controls, but they do not replace the two content-planning activities required before section drafting. The guide should be planned around the decisions and tasks its intended readers must perform.

NEW QUESTION: 29
You are designing a test strategy for a Claude-based pipeline that handles sensitive financial data.
Which two test types should be prioritized to cover both safety under attack and cross-component correctness? (Select two.)
A. Adversarial tests using prompt-injection and malformed-input cases.
B. Regression tests against a stable reference set of previously known-good outputs.
C. Smoke tests that verify core paths after each deployment.
D. Integration tests that verify end-to-end pipeline behavior across all components.
E. Unit tests targeting only individual prompt-template rendering logic.
Answer: (SHOW ANSWER)
Adversarial tests directly evaluate behavior under attack, including prompt injection, malformed records, malicious document content, credential-extraction attempts, and unauthorized tool requests. Integration tests provide the complementary coverage required by the question: they verify authentication, retrieval, prompt construction, model invocation, validation, logging, and downstream actions as one end-to-end system.
Regression, smoke, and unit testing remain valuable, but they do not jointly satisfy the two stated objectives.
A stable regression set measures previously established behavior; smoke tests provide shallow availability confirmation; and prompt-rendering unit tests cover only one component. Anthropic recommends explicit prompt-injection defenses and task-specific evaluations that represent both normal and edge-case traffic.
Prompt-injection mitigation

NEW QUESTION: 30
You are evaluating model-selection claims used by a peer team.
For each statement, select yes if the statement is generally true. Otherwise, select no.

Answer Area

Statements:	Select Yes or No:
Heavier-reasoning models typically consume more of the rate limit and run more slowly than lighter models on the same input.	Yes / No
A representative-sample evaluation on the target workload is the right basis for confirming a model selection.	Yes / No
The most capable model is always the right choice regardless of the workload profile.	Yes / No
Tiered routing can serve routine traffic with a lighter model while reserving a heavier model for traffic that needs its capability.	Yes / No
Selecting a model based on which version was most recently announced is the right basis for production decisions.	Yes / No

Answer:

Answer Area

Statements:	Select Yes or No:
Heavier-reasoning models typically consume more of the rate limit and run more slowly than lighter models on the same input.	Yes / No
A representative-sample evaluation on the target workload is the right basis for confirming a model selection.	Yes / No
The most capable model is always the right choice regardless of the workload profile.	Yes / No
Tiered routing can serve routine traffic with a lighter model while reserving a heavier model for traffic that needs its capability.	Yes / No
Selecting a model based on which version was most recently announced is the right basis for production decisions.	Yes / No

Explanation:
Yes, Yes, No, Yes, No
More capable reasoning configurations generally require greater computation, produce higher latency, and may consume larger token or rate-limit allocations than lighter alternatives. This makes the first statement generally accurate, although actual performance must be measured under the organization's specific provider and service tier.
Representative workload testing is essential because generic benchmark rankings do not establish performance on the organization's prompts, data, edge cases, tools, or required output format. Anthropic recommends benchmark tests specific to the use case, using actual prompts and data, followed by comparison of accuracy, quality, edge-case behavior, and cost. Choosing the Right Model The strongest model is not automatically the correct production choice. Model selection must balance capability, speed, cost, throughput, and operational constraints. Tiered routing implements this principle by serving predictable requests with an efficient model and escalating difficult cases to a more capable model.
Anthropic documents multi-model strategies in which lower-cost workers handle most traffic while harder decisions are escalated.
Recency is also not a valid selection criterion. A newly announced model must pass the workload's existing evaluation, security, latency, and cost gates before replacing a validated production model.
Study Guide references/topics: Model trade-offs; representative evaluations; tiered routing; escalation architectures; cost-capability optimization; model migration validation.

NEW QUESTION: 31
You are classifying chunking strategies by the corpus type each is best suited to.
For each chunking strategy, select the appropriate corpus type: "Long Structured Documents,"
"Heterogeneous Short Records," or "Code or Hierarchical Specifications."

Answer Area

Chunking Strategy:	Select the Appropriate Corpus Type:
function-level or section-level chunking for code modules	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
tree-aware chunking that follows code or specification hierarchy	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
per-record chunking where each record is one chunk	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
semantic chunking along clause or paragraph boundaries	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
fixed-size chunking with overlap for short records of similar length	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
hierarchical chunking that mirrors document section structure	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs

freecram.net

Answer:

Answer Area

Chunking Strategy:	Select the Appropriate Corpus Type:
function-level or section-level chunking for code modules	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
tree-aware chunking that follows code or specification hierarchy	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
per-record chunking where each record is one chunk	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
semantic chunking along clause or paragraph boundaries	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
fixed-size chunking with overlap for short records of similar length	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs
hierarchical chunking that mirrors document section structure	Long Structured Documents / Heterogeneous Short Records / Code or Hierarchical Specs

Explanation:
* Function-level or section-level chunking for code modules - Code or Hierarchical Specifications
* Tree-aware chunking that follows code or specification hierarchy - Code or Hierarchical Specifications
* Per-record chunking where each record is one chunk - Heterogeneous Short Records
* Semantic chunking along clause or paragraph boundaries - Long Structured Documents
* Fixed-size chunking with overlap for short records of similar length - Heterogeneous Short Records
* Hierarchical chunking that mirrors document section structure - Long Structured Documents Chunking must preserve the structural unit that carries meaning in the source corpus. Code and hierarchical specifications are best divided at function, module, class, or tree boundaries because arbitrary token cuts can separate definitions from their implementation or parent context. Heterogeneous short-record collections should generally preserve each record as an independent chunk. Fixed-size overlapping chunks are also effective when records have broadly similar lengths and lack meaningful internal hierarchy. Long structured documents benefit from semantic boundaries such as clauses and paragraphs, while hierarchical chunking preserves relationships among sections, subsections, and parent headings. Anthropic notes that chunk size, boundaries, and overlap materially affect retrieval performance; therefore, one universal chunking method is inappropriate. Anthropic Contextual Retrieval

Valid CCAR-P Dumps shared by ExamDiscuss.com for Helping Passing CCAR-P Exam! ExamDiscuss.com now offer the **newest CCAR-P exam dumps**, the ExamDiscuss.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com CCAR-P dumps with Test Engine here: https://www.examdiscuss.com/Anthropic/exam/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 32
You are responding to an adversarial input pattern in which users include text claiming admin authority and instructing the model to bypass safety restrictions.
Which combination of controls most effectively mitigates this attack pattern?
A. Trusting that the model will intrinsically recognize and reject all bypass attempts without prompt-level instructions, runtime classifiers, scoped permissions, or audit logging.
B. Prompt-level instructions that treat user content as untrusted data, runtime classifiers that detect override attempts, scoped tool permissions that cannot be elevated by user content, and audit logging of attempts.
C. Removing all safety restrictions and guardrails to eliminate the attack surface that bypass attempts target, accepting that this makes the assistant unrestricted for all inputs.
D. Granting users any privilege level they assert in their message content, on the assumption that cooperative behavior requires honoring self-declared authority without independent verification.
Answer: (SHOW ANSWER)
Self-declared administrative authority inside a prompt is untrusted data, not authenticated identity or authorization evidence. Option B correctly combines independent controls across the model, runtime, authorization, and monitoring layers.
Prompt instructions establish that user content cannot override system policy. Runtime classifiers detect known and generalized attempts to bypass controls. Tool permissions are enforced outside the model and must derive from authenticated identity, role, and approved scope-not statements contained in the conversation. Audit logging records the actor, attempted override, classifier result, tool requests, and final disposition for investigation and control improvement.
Anthropic recommends input screening, hardened system prompts, safe handling of untrusted content, narrowly scoped permissions, red-team testing, and continuous monitoring. Mitigate Jailbreaks and Prompt Injections Option A depends entirely on model behavior and provides no containment if the model fails. Option C removes the protections the attacker is attempting to defeat. Option D commits a fundamental authorization error by accepting an unverified claim as privilege elevation.
The strongest design also rate-limits repeated attempts, escalates suspicious activity, validates outputs, and requires human confirmation for consequential actions.
Study Guide references/topics: Direct prompt injection; untrusted user content; runtime classifiers; non-model authorization; scoped tools; audit logging; defense in depth.

NEW QUESTION: 33
A Claude architect needs to ensure that a security-hardening flag cannot be disabled by any individual engineer after it is set.
Which configuration scope correctly enforces this requirement?
A. User scope (~/.claude/settings.json) on each engineer's machine
B. Environment variable defined in the CI/CD pipeline only
C. Managed configuration applied centrally and marked as non-overridable
D. Project scope (.claude/settings.json) committed to the repository
Answer: (SHOW ANSWER)
Managed configuration is designed for organization-wide security and compliance controls that individual users and repositories must not override. Applying the hardening flag centrally establishes an administrative policy boundary and prevents engineers from disabling the control through user, project, or local settings.
Anthropic identifies managed scope as the appropriate location for security policies, non-overridable compliance requirements, and standardized configurations deployed by IT or DevOps. Managed values ordinarily take precedence over command-line arguments and every user-controlled settings scope. Claude Code Settings Option A depends on every engineer maintaining the setting and allows the user to edit or delete it. Option B affects only pipeline execution and does not protect local, interactive, IDE, or other execution paths. Option D standardizes the setting within the repository but does not make it tamper-resistant; contributors with repository write access could alter the file, use higher-precedence local settings where permitted, or operate outside the repository configuration.
The implementation should also verify active policy delivery, monitor configuration-change events, and test that startup or execution fails safely if the managed setting is absent or invalid. Central definition without enforcement and verification would not fully satisfy the requirement.
Study Guide references/topics: Managed settings; non-overridable controls; enterprise hardening; policy enforcement; configuration governance; defense against local override.

NEW QUESTION: 34
You are defining when to introduce a project subagent versus relying on Claude Code ' s general capabilities.
Which scenario most directly justifies a dedicated subagent?
A. The team has a recurring specialized task, such as database schema review, that requires a focused system prompt, narrow tool permissions, and a specific model selection across many sessions.
B. The team has a one-time ad hoc question that will not recur and does not require a focused system prompt, narrow tool permissions, or dedicated model selection.
C. The team has no recurring specialized tasks and uses Claude Code only for isolated general-purpose work that does not justify a dedicated system prompt or tool scope.
D. The team wants every Claude Code interaction to use the same generic system prompt with no task- specific specialization, narrow tool permissions, or dedicated model selection.
Answer: (SHOW ANSWER)
A dedicated subagent is justified when specialization is recurring and operationally valuable. Option A allows the team to encode a database-review role once, including its system prompt, triggering description, permitted tools, model, and review criteria. This improves consistency across sessions, isolates specialized context, and applies least privilege by withholding unrelated capabilities. Anthropic's custom subagent documentation supports task-specific prompts, model selection, tool allowlists, permission modes, hooks, skills, and persistent memory. A one-time question does not justify the maintenance overhead of a reusable agent definition, while Options C and D explicitly provide no specialization requirement. The subagent should also have clear invocation conditions, evaluation cases, ownership, and lifecycle review.
Study Guide references/topics: Claude Code subagents; task specialization; context isolation; model routing; scoped permissions; reusable automation.

NEW QUESTION: 35
You are designing a content moderation classifier that processes high volumes of user-generated comments under a tight per-message latency budget using well-defined classification labels.
Which model selection best aligns with the workload?
A. Opus, because every moderation decision requires maximum reasoning depth regardless of classification complexity.
B. Haiku, because its latency and cost profile align with high-volume classification workloads that require limited reasoning depth.
C. Sonnet, because larger general-purpose models are preferred even when workload latency requirements are strict.
D. Sonnet with extended thinking enabled, because deeper reasoning should be applied to every moderation request to improve edge-case handling.
Answer: B (LEAVE A REPLY)
Haiku is the appropriate starting point because the workload is high-volume, latency-sensitive, and based on a stable closed set of moderation labels. These characteristics favor a fast, cost-efficient model capable of consistent classification without incurring the additional inference time and expense associated with deeper reasoning.
Anthropic's model-selection guidance requires architects to balance capability, speed, and cost rather than automatically selecting the most capable model. Its content-moderation guidance specifically identifies Haiku as a cost-effective option for processing moderation workloads at substantial scale. Choosing the Right Model, Content Moderation Opus is disproportionate to a routine closed-set classification problem. Sonnet may become justified if evaluation demonstrates that Haiku fails materially on complex policy distinctions, multilingual ambiguity, or adversarial edge cases, but it should not be selected merely because it is larger. Enabling extended thinking on every request would further increase latency and token consumption without evidence that the additional reasoning improves the defined success metrics. The correct architectural practice is to establish a representative moderation evaluation set, validate Haiku against accuracy and safety thresholds, and escalate only the cases that genuinely need deeper reasoning.
Study Guide references/topics: Model selection; capability-latency-cost trade-offs; classification workloads; evaluation-driven routing; moderation architecture.

NEW QUESTION: 36
You are explaining the precedence of Claude Code configuration scopes to the team.
Which precedence ordering, from highest to lowest, is correct?
A. Local # managed # user # command-line arguments # project
B. Project # user # managed # local # command-line arguments
C. Managed # command-line arguments # local # project # user
D. User # project # local # command-line arguments # managed
Answer: C (LEAVE A REPLY)
The correct general precedence is managed configuration, command-line arguments, local configuration, project configuration, and then user configuration.

Managed configuration has the highest authority because it contains centrally administered organizational and compliance controls that ordinary engineers must not be able to override. Command-line arguments provide temporary session-level overrides where the relevant setting permits them. Local settings apply to one user within a specific repository and take precedence over the shared project configuration. Project settings apply to every contributor using that repository and override general user preferences. User settings form the lowest tier because they provide personal defaults across projects and apply only when no higher-precedence source supplies a value.
Anthropic documents this order explicitly: managed, command line, local, project, and user. It also notes that permission rules may merge rather than behave as simple scalar overrides, and certain security-sensitive keys have specialized precedence behavior. Claude Code Configuration Scopes The ordering allows personal customization without permitting it to defeat repository conventions or centrally mandated security requirements. Architects must still examine the merge behavior of the individual configuration property rather than assuming every array or permission rule is replaced wholesale.
Study Guide references/topics: Claude Code configuration; scope precedence; managed policies; local overrides; project settings; user defaults.

NEW QUESTION: 37
You are configuring tool permissions for a Claude-based assistant. The assistant's defined responsibilities require read access to a knowledge base and write access to a draft queue, and nothing else.
Which scoping design best applies least privilege?
A. Allow access to all tools with read permissions globally and restrict write permissions to the draft- queue tool, without scoping to only the specific tools required for defined tasks.
B. Per-role allow-list of exactly the read-knowledge-base and write-draft-queue tools, enforced at the orchestration layer.
C. Allow access to every tool in the catalog and rely on the model to decline tools it should not use.
D. Disable all tools entirely to achieve a minimal permission surface, accepting that the assistant can no longer perform the read-knowledge-base or write-draft-queue tasks it was designed for.
Answer: (SHOW ANSWER)
Least privilege means granting the exact capabilities required for the defined responsibility-no more and no less. Option B permits the assistant to read the approved knowledge base and write only to the draft queue.
Enforcing the allow-list at the orchestration layer creates a deterministic authorization boundary outside the model's prompt and reasoning.
The rule should also be role-specific so different users or operational roles receive only the tools and data scopes they require. The knowledge-base tool should remain read-only, while the draft-queue tool should be restricted to draft creation rather than external sending, approval, or production publication.
Option A exposes every globally readable tool, creating unnecessary access to unrelated information. Option C relies on voluntary model behavior instead of enforceable authorization. A mistaken tool selection or successful prompt injection could therefore invoke an inappropriate capability. Option D minimizes permissions by eliminating the assistant's required functionality and is not a viable least-privilege implementation.
Anthropic's permission guidance distinguishes allow, ask, and deny rules and confirms that these controls are enforced by Claude Code rather than by model instructions. Configure Permissions Study Guide references/topics: Least privilege; per-role authorization; orchestration-layer enforcement; capability scoping; read/write separation; tool allow-lists.

NEW QUESTION: 38
You are listing characteristics of robust guardrail design for an enterprise deployment.
Which two characteristics belong on the list? (Select two.)
Each correct answer presents a complete solution.
A. Centralized log retention for guardrail violations with quarterly review by the security team.
B. Per-role tool allow-lists enforced at the orchestration layer before any tool call executes.
C. User feedback channels that route reported guardrail failures into the product backlog for triage.
D. Periodic refresh of the system prompt wording to keep refusal language current and clear.
E. Adversarial-input coverage in the evaluation set with regression tracking on guardrail performance.
Answer: (SHOW ANSWER)
Robust guardrails require both preventive enforcement and continuous empirical validation. Per-role tool allow-lists satisfy the preventive requirement by restricting each authenticated role to explicitly authorized tools. Enforcement must occur in the orchestration or permission layer before execution; relying on the model to decide whether a tool call is authorized is not an adequate security boundary. Anthropic recommends allow- list-based validation and provides permission rules and pre-tool-use controls capable of denying calls before they execute.
Adversarial evaluation with regression tracking supplies the validation component. The evaluation set should include prompt injections, indirect injections, encoded instructions, multi-turn escalation, role-confusion attempts, and requests designed to trigger unauthorized tool use. Guardrail performance must then be measured across releases so that changes to prompts, models, tools, or orchestration logic do not silently reduce protection.

Centralized logging and user feedback are useful detective and lifecycle controls, but neither prevents an unsafe action nor proves guardrail effectiveness. Periodically refreshing refusal wording is also insufficient unless the revised behavior is evaluated against defined safety criteria. The strongest design therefore combines deterministic authorization with adversarial regression testing.
Study Guide references/topics: Claude Code permissions and pre-tool enforcement; tool allow-list guidance; adversarial evaluations; regression testing; least-privilege orchestration.

NEW QUESTION: 39
You are defining transparency practices for a customer-facing assistant whose responses are materially shaped by AI.
Which transparency practice most directly supports responsible deployment?
A. Misrepresent the AI's role in producing responses in order to make the assistant feel more trustworthy or more human, undermining informed user consent and organizational transparency.
B. Refuse to answer any user question about how the responses were produced or whether AI was involved, treating the AI's role as confidential operational information.
C. Disclose AI involvement to end users in line with the organization's transparency policy and provide a documented path to reach a human when needed.
D. Disclose AI involvement only to internal staff and operators while withholding that information from the end users whose interactions are materially shaped by the AI system.
Answer: (SHOW ANSWER)
Responsible deployment requires users to receive an accurate representation of the system with which they are interacting. Option C provides that transparency while also establishing a practical escalation path.
Disclosure should be proportionate to the material role AI plays, expressed in language appropriate to the audience, and aligned with organizational policy and applicable regulatory obligations.
A documented human-contact route is important when the assistant cannot resolve an issue, produces a disputed result, encounters a high-impact exception, or handles a matter requiring human authority. Anthropic' s Responsible Scaling Policy recognizes escalation to human reviewers as an appropriate safeguard for edge cases and situations requiring human judgment. Responsible Scaling Policy Options A and D intentionally conceal material AI involvement from affected users, weakening informed decision-making, auditability, and trust. Option B is also unsuitable because transparency does not require revealing proprietary prompts, confidential controls, or internal security mechanisms; however, it does require an honest explanation of AI involvement and the system's operational role.
The correct design combines disclosure, understandable limitations, human escalation, and records demonstrating that the organization's transparency requirements are consistently applied.
Study Guide references/topics: AI transparency; informed user interaction; human escalation; responsible deployment; disclosure policy; operational accountability.

NEW QUESTION: 40
You are presenting an architectural decision to a mixed audience that includes an executive sponsor and the engineering leads who will implement the decision.
Which presentation strategy best serves both audiences?
Freecram.net
A. Open with a deep dive into low-level implementation details targeted at engineering leads, and stop there without addressing the business outcomes or trade-offs the executive sponsor needs.
B. Lead with the decision, the business outcomes it serves, and the trade-offs accepted; follow with the technical rationale, alternatives, and implementation implications for the engineering audience.
C. Skip the rationale, alternatives, and trade-off discussion entirely and simply announce the chosen decision, leaving both audiences without the context needed to implement or validate it.
D. Present a single undifferentiated narrative that addresses technical and business concerns with equal weight throughout, treating both audiences as requiring the same depth on every section.
Answer: (SHOW ANSWER)
Option B uses progressive disclosure to meet the decision needs of both audiences. The executive sponsor first receives the selected direction, intended business outcomes, major constraints, cost or performance implications, and accepted trade-offs. This establishes why the decision matters and whether it remains aligned with strategic objectives.
The presentation can then move into the technical rationale required by engineering leads: component responsibilities, integration contracts, model and retrieval choices, security controls, rejected alternatives, implementation dependencies, evaluation gates, operational consequences, and migration requirements. This structure gives engineers enough evidence to implement the decision without forcing the executive audience to interpret unnecessary low-level detail.
Option A excludes the sponsor's concerns and fails to connect technical design to business value. Option C removes the context required for governance and implementation, making the decision appear arbitrary.
Option D assumes that all participants require identical information density, which usually produces a presentation that is too technical for executives and insufficiently precise for engineers.
The strongest communication pattern is layered but internally consistent: one decision, one rationale, and multiple levels of detail tailored to the responsibilities of the audience.
Study Guide references/topics: Mixed-audience communication; progressive disclosure; business outcomes; architectural trade-offs; technical rationale; implementation implications.

NEW QUESTION: 41

You are building an ethics review checklist for deployments supported by artificial intelligence (AI). Which two checks belong on the list? (Select two.) Each correct answer presents a complete solution.

A. Confirm that vendor licensing terms permit the planned production use of the model.

B. Confirm that high-impact decisions retain human accountability rather than being attributed to the model.

C. Restrict the ethics review to outputs that exceed a defined model-confidence threshold.

D. Confirm that latency and throughput targets are met across the supported user populations.

E. Verify that outputs do not rely on generalizations about people that the underlying data does not support.

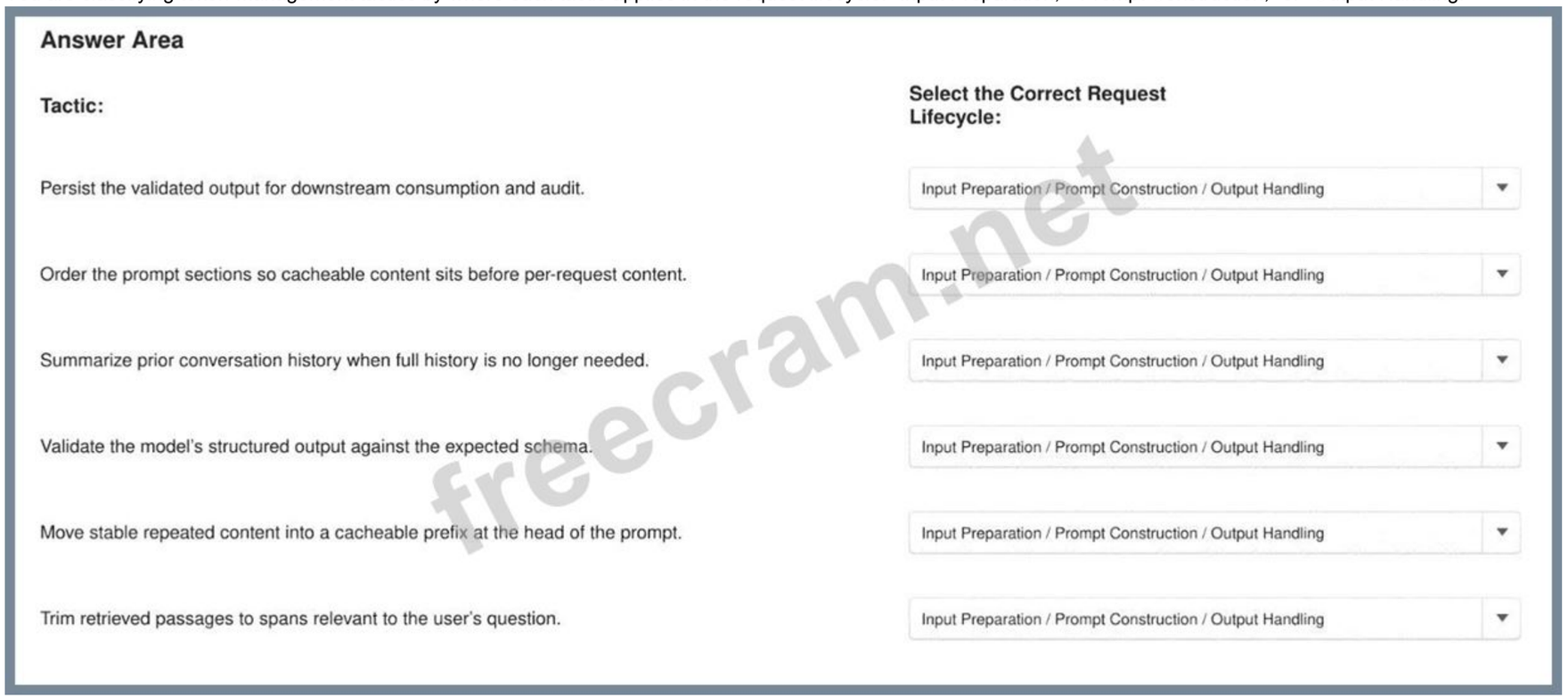
Answer: (SHOW ANSWER)

High-impact AI-supported decisions must retain clearly assigned human accountability. Claude may provide recommendations, classifications, or generated content, but responsibility for consequential outcomes cannot be transferred to the model. An ethics review should also investigate whether outputs make unsupported generalizations about individuals or groups, because such behavior can produce discriminatory, misleading, or harmful outcomes. Therefore, B and E directly address ethical accountability and fairness. Licensing in option A is an important legal and procurement concern, but it is not the strongest ethics-focused control presented.

Latency and throughput in option D are operational performance requirements. Option C is inappropriate because ethics risks are not limited to outputs above a confidence threshold; confidence values may also be poorly calibrated and do not prove that an output is fair, accurate, or ethically acceptable.

NEW QUESTION: 42

You are classifying token-management tactics by where each tactic applies in the request lifecycle: "Input Preparation," "Prompt Construction," or "Output Handling."



freecram.net

Answer:

Answer Area

Tactic:	Select the Correct Request Lifecycle:
Persist the validated output for downstream consumption and audit.	Input Preparation / Prompt Construction / Output Handling
Order the prompt sections so cacheable content sits before per-request content.	Input Preparation / Prompt Construction / Output Handling
Summarize prior conversation history when full history is no longer needed.	Input Preparation / Prompt Construction / Output Handling
Validate the model's structured output against the expected schema.	Input Preparation / Prompt Construction / Output Handling
Move stable repeated content into a cacheable prefix at the head of the prompt.	Input Preparation / Prompt Construction / Output Handling
Trim retrieved passages to spans relevant to the user's question.	Input Preparation / Prompt Construction / Output Handling

Explanation:
* Persist the validated output for downstream consumption and audit - Output Handling
* Order prompt sections so cacheable content appears before per-request content - Prompt Construction
* Summarize prior conversation history when full history is no longer needed - Input Preparation
* Validate the model's structured output against the expected schema - Output Handling
* Move stable repeated content into a cacheable prefix at the beginning of the prompt - Prompt Construction
* Trim retrieved passages to spans relevant to the user's question - Input Preparation Input preparation determines what evidence and history should enter the request, so summarization and retrieval trimming belong there. Prompt construction determines ordering and cache boundaries; stable repeated instructions must precede dynamic per-request content to maximize cache reuse. Output handling begins after generation and includes schema validation, persistence, auditing, and downstream delivery.
Mixing these responsibilities creates inefficient prompts and weak validation boundaries. Anthropic's context guidance emphasizes curating the smallest useful context because unnecessary tokens can reduce recall and accuracy. Prompt caching similarly depends on a stable reusable prefix, while structured outputs or schema validation protect downstream systems from malformed responses. Context windows; prompt caching

NEW QUESTION: 43
You are running a risk assessment on a planned Claude-based deployment and must complete the inventory steps before assessing threats against assets.

Which two steps must be completed BEFORE assessing threats against assets to estimate likelihood and impact? (Select two.) Each correct answer presents part of the solution.
A. Document the assessment outcome with risks, mitigations, residual risk, and acceptance owners.
B. Recommend mitigations and residual-risk acceptance for the risks that remain after analysis.
C. Identify the assets that the deployment touches, along with the sensitivity of each asset.
D. Recommend mitigations and residual-risk acceptance for risks that remain.
E. Enumerate the threat actors and attack vectors relevant to the deployment.
Answer: (SHOW ANSWER)
Threat assessment requires an inventory of what must be protected and who or what may threaten it. Asset identification should cover prompts, retrieved documents, credentials, model outputs, audit logs, tools, external systems, and sensitive business or personal data. Sensitivity classifications establish the potential confidentiality, integrity, safety, and availability impact of compromise. Threat actors and attack vectors must then be enumerated so likelihood and impact can be assessed against specific assets. Mitigation recommendations, residual-risk acceptance, ownership, and final documentation follow the analysis; they cannot be completed coherently before the threats are understood. Anthropic's security transparency materials similarly emphasize maintaining threat models that reflect relevant attacker tactics and techniques. Anthropic Transparency Hub

NEW QUESTION: 44
You are compiling factors that should drive the choice between Model Context Protocol (MCP), direct API integration, and agent-to-agent handoff.
Which two factors belong on the list? (Select two.)
Each correct answer presents a complete solution.
A. Whether the vendor publishes detailed reference documentation for each candidate protocol.
B. Whether the team has prior implementation experience with any of the candidate protocols.
C. Whether the underlying transport supports encryption in transit between the integrated services.
D. Whether the integration must be portable across multiple AI clients in the ecosystem.
E. Whether the interaction is stateless and latency-sensitive or stateful and longer-running.
Answer: (SHOW ANSWER)
The interaction model and portability requirement are architectural decision drivers. MCP is appropriate when tools or data sources must be exposed through a standardized interface reusable across compatible AI clients.
Direct API integration is often preferable for a narrow, stateless, latency-sensitive call where the application already owns the integration logic. Agent-to-agent handoff is better suited to stateful or longer-running delegation between independently responsible agents. Documentation quality and team familiarity influence implementation effort but do not determine the correct integration pattern. Encryption in transit is a mandatory security property for production integration rather than a differentiating architectural requirement; each viable approach must provide it. Anthropic distinguishes direct Messages access from stateful managed- agent infrastructure, while MCP supplies a standardized tool integration layer. Claude Platform documentation

NEW QUESTION: 45
A document analysis service processes legal filings averaging 80,000 tokens each. Each filing is queried by attorneys an average of 14 times during a case. The current architecture sends the full filing on every query.
The CFO has asked you to reduce per-query costs while preserving response quality. The security officer requires that filing contents not be stored outside Fabrikam ' s tenancy.
Which optimization approach should you recommend?
A. Summarize each filing once at intake and run all subsequent queries against the summary.
B. Cache the filing as the prompt prefix for reuse across the 14 queries per case.
C. Index filings in a vector store and retrieve only the relevant passages per query.
D. Move the workload to a smaller Claude model to reduce the per-token cost paid.
Answer: (SHOW ANSWER)
The 80,000-token filing is a large, stable prefix reused across fourteen requests, making prompt caching the strongest optimization. The first request creates the cache; subsequent identical-prefix requests use lower- priced cache reads while retaining the complete filing, preserving the evidence available to Claude. Anthropic documents cache reads at a fraction of standard input cost and confirms that caches are isolated by organization and, on supported platforms, by workspace. Cached representations remain in memory rather than persistent storage. Options A and C may omit context through summarization or retrieval and therefore require additional quality evaluation. Option D lowers unit cost but may reduce reasoning quality. The design must preserve prefix identity, choose an appropriate TTL, and verify the tenancy configuration.

Study Guide references/topics: Prompt caching; cache isolation; stable prefixes; long-document analysis; cost optimization.

NEW QUESTION: 46
A healthcare organization is evaluating two Claude-powered AI architectures for a clinical documentation assistant. Architecture X produces higher output quality scores but costs $0.18 per documentation session and averages 4.2 seconds per response. Architecture Y scores slightly lower on quality metrics but costs $0.09 per session and averages 2.1 seconds per response. The stated SLA requires responses under 3 seconds, and the annual volume is projected at 2 million documentation sessions.
Which evaluation approach correctly applies business value pillar analysis to this decision?
A. Select Architecture Y based solely on the 50% cost reduction, since solution cost is the most important value pillar in healthcare budget-constrained environments.
B. Select Architecture X because the higher quality scores justify the cost premium, and any SLA gap can be addressed through infrastructure optimization after deployment.
C. Eliminate Architecture X on SLA grounds, then evaluate Architecture Y against the efficiency and solution cost pillars by calculating annual cost difference and assessing whether the quality delta materially impacts clinical workflow productivity.
D. Recommend a hybrid approach using Architecture X for complex cases and Architecture Y for routine cases, without additional analysis, since this preserves quality where it matters most.
Answer: (SHOW ANSWER)
Architecture X fails a stated production gate: its 4.2-second average exceeds the required three-second response SLA. It should therefore not remain a qualifying architecture unless it is redesigned and retested before selection. A promise to optimize after deployment is not evidence of compliance.
Architecture Y meets the latency requirement at 2.1 seconds. Its annual inference cost is $180,000 at two million sessions, compared with $360,000 for Architecture X-a difference of $180,000 annually. However, cost alone does not establish suitability. The organization must determine whether Y's lower quality score causes material omissions, correction effort, clinician dissatisfaction, documentation risk, or reduced workflow productivity.
Option C correctly treats the SLA as a threshold and then analyzes the remaining architecture across cost, efficiency, quality, and productivity. Option A ignores clinical quality. Option B disregards an explicit SLA failure. Option D may eventually be viable, but it introduces routing complexity and leaves unanswered how complex cases are detected, whether X's latency becomes acceptable for them, and how both paths are evaluated.
The final choice should be gated by clinical-quality evaluation, human review requirements, safety metrics, operational latency, and total annual cost.
Study Guide references/topics: Business-value pillars; performance gates; solution cost; workflow efficiency; quality trade-offs; healthcare architecture evaluation.

Valid CCAR-P Dumps shared by ExamDiscuss.com for Helping Passing CCAR-P Exam! ExamDiscuss.com now offer the **newest CCAR-P exam dumps**, the ExamDiscuss.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com CCAR-P dumps with Test Engine here: https://www.examdiscuss.com/Anthropic/exam/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 47
You are evaluating prompting claims in a peer's design document.
For each claim, select yes if the claim reflects sound practice. Otherwise, select no.

Answer Area

Claim:	Select Yes or No:
Zero-shot is appropriate as a starting point for closed-set classification when categories and descriptions are provided in the prompt.	Yes / No
Adding a few-shot example set with exact field names anchors output formatting on extraction tasks.	Yes / No
Chain-of-thought should be the default on every task regardless of complexity.	Yes / No
Leading-question framing is best mitigated by asking for an even-handed comparison with stated criteria.	Yes / No
Behavioral differences across model versions can be ignored once a prompt has been validated on one version.	Yes / No

Answer:

Answer Area

Claim:	Select Yes or No:
Zero-shot is appropriate as a starting point for closed-set classification when categories and descriptions are provided in the prompt.	Yes / No
Adding a few-shot example set with exact field names anchors output formatting on extraction tasks.	Yes / No
Chain-of-thought should be the default on every task regardless of complexity.	Yes / No
Leading-question framing is best mitigated by asking for an even-handed comparison with stated criteria.	Yes / No
Behavioral differences across model versions can be ignored once a prompt has been validated on one version.	Yes / No

freecram.net

Explanation:
Yes, Yes, No, Yes, No
Zero-shot prompting is a reasonable baseline for a clearly defined closed-set classifier because the model receives both the permissible labels and their meanings. Performance should then be measured against an evaluation set before additional prompt complexity is introduced. Few-shot examples are also sound for extraction because examples containing the exact field names demonstrate the required schema and reduce formatting ambiguity. Anthropic identifies examples as one of the most reliable methods for steering output format, structure, and consistency. Prompting Best Practices Chain-of-thought is not an unconditional default. Additional reasoning is appropriate when the task requires substantive decomposition or complex judgment, but routine classification and extraction may gain little while consuming more tokens and increasing latency. The technique should be selected according to evaluated task requirements. An even-handed comparison governed by explicit criteria mitigates leading-question bias by requiring symmetrical treatment of alternatives rather than presupposing a preferred conclusion. Finally, prompt validation is model-specific. New model versions may follow instructions differently, change formatting behavior, alter tool selection, or produce different latency and cost characteristics. Anthropic's migration guidance consequently recommends reevaluating prompts and workloads against the new baseline. Model Migration Guide Study Guide references/topics: Zero-shot prompting; few-shot examples; reasoning techniques; neutral framing; model migration and regression evaluation.

NEW QUESTION: 48
You are an architect supporting the iteration phase of a deployed Claude-based system.
Which activity most directly fits this phase?

A. Rebuild the entire system architecture from scratch at the start of every iteration cycle regardless of what production telemetry, evaluations, and stakeholder feedback indicate is needed.
B. Stop measuring production outcomes once the deployment has successfully launched, treating the go- live milestone as the end of the evaluation and iteration cycle.
C. Discard the evaluation framework and reference set once the deployment is in production, accepting that future iterations will have no structured basis for measuring the impact of changes.
D. Review production telemetry, sampled output evaluations, and stakeholder feedback to identify the highest-impact change for the next cycle, then plan the change against the evaluation framework.
Answer: (SHOW ANSWER)
Iteration is an evidence-driven lifecycle phase. Option D combines three essential inputs: production telemetry, sampled output evaluation, and stakeholder feedback. Telemetry exposes latency, cost, errors, tool failures, retrieval behavior, and operational anomalies. Output sampling identifies quality, safety, grounding, or consistency failures that infrastructure metrics may not reveal. Stakeholder feedback establishes whether the system continues to satisfy real business and user needs.
The architect should prioritize the highest-impact validated problem, define the intended improvement, and test the proposed change against the existing evaluation framework and reference set. Anthropic emphasizes that evaluations make behavioral changes visible before they affect users and that their value compounds throughout an agent's lifecycle. Demystifying Evals for AI Agents Option A replaces targeted iteration with unnecessary redesign. Option B incorrectly treats deployment as the end of measurement. Option C removes the baseline needed to distinguish genuine improvement from regression or noise.
A mature iteration cycle preserves evaluation continuity, adds production failures as new test cases, validates changes offline, conducts controlled rollout, monitors results, and updates documentation and known limitations.
Study Guide references/topics: Production iteration; telemetry; sampled evaluation; stakeholder feedback; regression testing; evaluation continuity; change prioritization.

NEW QUESTION: 49
A team manager wants all engineers working on the same repository to share identical MCP server definitions without manual synchronization.
Which configuration approach satisfies this requirement?
A. managed configuration pushed to all endpoints by the administrator
B. environment variables set at the operating-system level on each workstation
C. project-scope .claude/settings.json and .mcp.json files committed to the repository
D. each engineer maintains a personal ~/.claude/settings.json with the shared definitions
Answer: (SHOW ANSWER)
Project-scoped configuration is the appropriate mechanism for repository-specific settings that must remain consistent across collaborators. Claude Code recognizes `.claude/settings.json` as the shared project settings file and `.mcp.json` as the project-level MCP server configuration. When these files are committed to source control, every engineer receives the same definitions through the repository's normal clone, pull, review, and versioning workflow.
Anthropic's scope hierarchy distinguishes project configuration from user and local configuration. Files under the user scope apply to one engineer and are not automatically shared. Operating-system environment variables are also workstation-specific and require separate configuration management. Managed settings can be distributed centrally, but they are primarily intended for organization-wide security policies and controls that must be enforced by administrators; they are less appropriate when the configuration belongs specifically to one repository.
Shared configuration should contain server definitions and environment-variable references, but not plaintext credentials. Each engineer can resolve required secrets through an approved local secret mechanism while retaining identical server names, transports, arguments, and non-sensitive settings. Project MCP servers also remain subject to workspace trust and approval controls before execution.
Study Guide references/topics: [Claude Code configuration scopes](https://docs.anthropic.com/en/docs
/claude-code/settings); project-scoped MCP servers; `.mcp.json`; repository-based operational standardization; secure secret separation.

NEW QUESTION: 50
You are a platform architect designing an internal Claude-based assistant that serves both finance analysts and external auditors. Each population must access only documents permitted by its role.
Where should role-based access control be enforced in the pipeline?
Freecram.net
A. Inside the system prompt as a natural-language instruction for Claude to ignore unauthorized documents.
B. At the retrieval layer, before any role-restricted content reaches the prompt-construction step or the model.
C. Nowhere in the pipeline; rely on the model's general refusal behavior to reject unauthorized document access without any enforced access control.
D. After the response is generated, by post-filtering content that should not have been retrieved.
Answer: (SHOW ANSWER)

Authorization must be enforced before restricted information enters model context. The retrieval service should authenticate the user, resolve role and document permissions, and filter candidate records before prompt construction. This prevents unauthorized content from appearing in prompts, model traces, cached context, tool results, or generated responses. A system-prompt instruction is behavioral guidance, not an enforceable access-control boundary. Post-generation filtering is also inadequate because the model has already received and processed the restricted data, creating exposure through output, logging, or indirect inference. Relying on general refusal behavior provides no deterministic authorization. The correct architecture treats RBAC as an application and data-layer control, with the model receiving only information the requesting principal is entitled to access.
Study Guide references/topics: Retrieval-layer RBAC; authorization before inference; least privilege; sensitive-context isolation.

NEW QUESTION: 51
Engineering leadership wants to roll out Claude Skills to 280 developers across 14 teams. Skills will encode internal coding standards, code-review checklists, and incident-postmortem templates. Leadership has asked how to govern Skill authorship so that Skills remain trustworthy without bottlenecking on a single central team.
Which governance model should you recommend?
A. Per-developer authorship across the 280 engineers with no team-level coordination required.
B. Centralized authorship by a single platform team responsible for every Skill produced.
C. Fully decentralized authorship across the 14 teams with no review before publication.
D. Federated authorship across the 14 teams with a central review and publication gate.
Answer: (SHOW ANSWER)
Federated authorship allows each domain team to maintain the procedural knowledge it understands while preserving organization-wide trust through a central publication gate. Team authors can update coding standards, review practices, and postmortem procedures without waiting for a single platform group to write every change. Central reviewers can validate ownership, security, tool permissions, duplication, versioning, testing evidence, and compliance before publication. Per-developer or fully decentralized publication creates inconsistent and potentially unsafe Skills. Fully centralized authorship becomes a scalability bottleneck and separates content maintenance from domain expertise. Anthropic describes Skills as packages of instructions, scripts, and resources and warns that repository Skills may grant broad tool access, making formal review essential. Claude Code Skills

NEW QUESTION: 52
You are designing a feedback session for a deployment in flight.
Which structure best supports productive stakeholder feedback?
A. Hold an open-ended meeting with no pre-distributed agenda or artifacts, and rely on participants' memory to carry decisions and follow-up owners forward.
B. Define a focused agenda, share the artifacts in advance, capture decisions and follow-ups in writing, and confirm action owners and dates.
C. Distribute artifacts at the session start rather than in advance, so stakeholders review materials in real time without preparation before the discussion begins.
D. End the session without recording decisions, follow-up items, action owners, or dates, relying on participant memory to carry the session's outcomes forward.
Answer: (SHOW ANSWER)
A structured feedback session must enable informed review, traceable decisions, and accountable follow- through. Option B allows stakeholders to examine architecture diagrams, evaluation results, risk registers, or prototypes before the meeting. The focused agenda keeps discussion aligned with decisions that must be made rather than allowing the session to become an unstructured status conversation. Written decision records prevent different participants from leaving with incompatible interpretations. Assigning an owner and due date to every action converts feedback into executable delivery work. Options A, C, and D depend heavily on memory, provide inadequate preparation time, and create weak decision traceability. In-flight deployments particularly require documented feedback because accepted changes may affect scope, cost, risk, testing, and release timing.
Study Guide references/topics: Stakeholder feedback; decision records; action ownership; lifecycle alignment; communication planning.

NEW QUESTION: 53
A Claude architect is configuring a shared Claude Code environment for a twelve-person development team.
Which two configuration decisions most directly establish a consistent, secure team environment? (Select two.)
A. Commit shared MCP server definitions and tool permissions in project-scope configuration files alongside the repository.
B. Apply managed configuration to enforce non-overridable security and compliance policies across all team members.
C. Set editor theme and display preferences at the project scope so they are uniform across workstations.
D. Instruct each engineer to configure their preferred Claude model in personal user-scope settings.

E. Store the team's shared API key in the project-scope settings.json so all engineers use the same credential.
Answer: (SHOW ANSWER)
Project-scope configuration is appropriate for shared MCP definitions, permissions, hooks, and other repository-specific behavior that every team member should receive consistently. Managed configuration complements this by enforcing organization-wide security and compliance requirements that individual developers cannot override. Together, these mechanisms separate collaborative project configuration from centrally controlled policy. Editor themes are personal preferences and do not materially establish a secure environment. Allowing each engineer to select an arbitrary model can undermine consistency when model choice affects behavior, cost, or approved-use requirements. Storing a shared API key in a committed project file is a serious credential-management failure. Claude Code documentation identifies .claude/settings.json as the shared project file and managed settings as the highest-precedence organizational policy layer. Claude Code: Settings files and precedence

NEW QUESTION: 54
You are selecting a pattern for a compliance Q & A assistant that must answer policy questions with citations to the authoritative internal source set. Latency, cost, and audit predictability are prioritized.
Which pattern is the best fit?
A. Augmented LLM with retrieval-augmented generation over the indexed authoritative corpus and citation rendering on each answer.
B. Multi-agent orchestration with a planner, researcher, and writer agent for every query.
C. A pure agent loop with open web-browsing tools to surface the most current policy information, without constraining retrieval to the authoritative internal corpus.
D. A static prompt with the entire policy corpus concatenated into every request.
Answer: (SHOW ANSWER)
A constrained retrieval-augmented generation architecture best satisfies the grounding, citation, latency, cost, and audit requirements. The system first indexes the approved internal policy corpus. At request time, it retrieves a small set of relevant passages, supplies those passages to Claude with stable source identifiers, and requires the answer to cite the supporting material. Anthropic's search-result content blocks are specifically designed for this pattern: they allow custom RAG systems to provide source-attributed internal content and enable Claude to render citations against those results.
The retrieval boundary is critical. Only approved and version-controlled policy sources should be searchable, with document identifiers, effective dates, access controls, retrieval scores, and cited passages retained for audit reconstruction. This produces substantially more predictable evidence than unconstrained web research.
A planner-researcher-writer system introduces unnecessary model calls, orchestration complexity, cost, latency, and behavioral variance for routine policy lookup. Open-web browsing violates the authoritative- source requirement and increases prompt-injection and provenance risk. Concatenating the complete corpus into every prompt wastes tokens, increases latency, and can reduce retrieval precision as the corpus expands.
Study Guide references/topics: Search-result blocks for cited RAG; Claude citations; authoritative-corpus retrieval; source attribution; auditable context engineering.

NEW QUESTION: 55
You are identifying signals that a deployment should re-enter design rather than continue iterating in place.
Which signal most directly indicates the need for a new design cycle?
A. A runbook step requires a clarification edit to improve on-call guidance accuracy, which can be handled as a documentation update without changes to component responsibilities or core contracts.
B. A dashboard alert threshold needs a small numerical adjustment to reduce false-positive noise, which can be handled as an operational configuration change without a new design cycle.
C. A minor copy edit is requested in a customer-facing string within the existing UI, which can be handled as a localized content change without altering component responsibilities or contracts.
D. The system's current architecture cannot meet the new requirements without changes to component responsibilities or core contracts.
Answer: D (LEAVE A REPLY)
A new design cycle is warranted when the required change crosses an architectural boundary. Option D explicitly states that the existing architecture cannot satisfy the new requirements without redistributing component responsibilities or altering core contracts. Such changes may affect service ownership, orchestration, data flows, integration interfaces, security boundaries, failure handling, evaluation strategy, and operational accountability. They therefore require renewed discovery, impact analysis, design review, stakeholder approval, and regression planning.
Options A, B, and C are localized maintenance activities. Clarifying a runbook improves operational documentation but does not change the deployed system's structure. Adjusting an alert threshold is a controlled operational tuning activity, provided the change remains within established monitoring policy.
Editing customer-facing copy is similarly confined to the presentation or content layer. None of these changes inherently invalidates component contracts or architectural assumptions.
An architect should distinguish between iteration within an approved design and evidence that the design itself no longer supports the required outcome. Re-entering design for every minor adjustment creates unnecessary governance overhead; continuing local iteration after architectural assumptions have failed creates unmanaged technical and operational risk. Option D is the only signal that establishes a structural incompatibility.

Study Guide references/topics: Lifecycle feedback loops; design re-entry criteria; architectural significance; component responsibilities; interface contracts; controlled operational iteration.

NEW QUESTION: 56
You are choosing the level of detail for an implementation guide. The audience is a delivery team that will build the deployment.
Which guidance composition best serves them?
A. Component responsibilities, contracts between components, sequence diagrams of the dominant flows, configuration parameters with defaults, and operational runbooks.
B. Component responsibilities and interface contracts only, without sequence diagrams of the dominant flows, configuration parameters with defaults, or runbooks to guide operational tasks.
C. An architecture overview and sequence diagrams for the dominant flows, without interface contracts, configuration-parameter tables, or operational runbooks for the delivery team to follow.
D. An architecture overview and a list of known limitations, without component-level diagrams, interface contracts, configuration parameters, or operational runbooks to support implementation.
Answer: (SHOW ANSWER)
A delivery team needs implementation-level direction rather than an executive architecture summary.
Component responsibilities establish ownership and boundaries, while interface contracts define schemas, accepted inputs, outputs, errors, and behavioral expectations. Sequence diagrams show how components interact across dominant and failure flows. Configuration tables with defaults prevent inconsistent assumptions between environments. Operational runbooks ensure that deployment, recovery, monitoring, and routine maintenance are addressed during implementation rather than improvised after release. Options B, C, and D omit one or more elements required to translate architectural intent into a buildable and supportable system. The guide should enable normal delivery work without continuous dependence on undocumented knowledge held by the architect.
Study Guide references/topics: Implementation documentation; component contracts; sequence diagrams; configuration management; operational readiness.

Valid CCAR-P Dumps shared by ExamDiscuss.com for Helping Passing CCAR-P Exam! ExamDiscuss.com now offer the **newest CCAR-P exam dumps**, the ExamDiscuss.com CCAR-P exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com CCAR-P dumps with Test Engine here: https://www.examdiscuss.com/Anthropic/exam/CCAR-P/premium/ (**130** Q&As Dumps, **35%OFF** **Special Discount Code: freecram**)