Claude Certified Associate - Foundations (CCAO-F) Questions and Answers

25 practice questions with answers and explanations, covering every exam domain. Taken from our bank of 150 reviewed questions.

Looking for exam dumps? Use these original practice questions instead. These are original practice questions written by CloudYeti. They are NOT real exam questions and are not affiliated with or endorsed by Anthropic.

Take the full timed practice test free:
https://learn.cloudyeti.io/free/claude-certified-associate-foundations-practice-test

DOMAIN 1 / PROMPTING AND TASK EXECUTION

1. An operations analyst asks Claude to turn weekly incident notes into an executive update. The first result is polished but omits owners and deadlines. What is the best next step?

A) Switch to a longer conversation so Claude can infer the missing details

B) Raise the temperature so the response explores more possibilities

C) Ask Claude to make the update better

D) Provide a required structure with sections for impact, owner, deadline, evidence, and open decisions

Answer: D. A concrete output structure and explicit required fields make success testable. More context or creativity does not supply missing requirements, and Claude should not invent owners or deadlines.

2. A recruiting team needs candidate summaries in a house style that is difficult to describe but has three excellent approved examples. Which prompt design is most effective?

A) Ask Claude to explain its hidden reasoning before writing

B) Attach the examples and ask Claude to follow their structure and tone while using only facts from the new resume

C) Ask for a professional summary without examples

D) Paste every summary the team has ever written

Answer: B. A small set of representative examples plus a factuality constraint teaches the target pattern without flooding the context. Vague instructions and excessive examples are less reliable.

3. A manager wants Claude to compare twelve vendor proposals and recommend a shortlist. The proposals use different formats. What workflow best reduces missed requirements?

A) Ask for the winner in one prompt

B) Have Claude extract each proposal into the same criteria table, verify citations, then compare the normalized rows

C) Ask Claude to rank vendors by writing quality

D) Summarize all proposals into one paragraph before comparing them

Answer: B. Decomposing extraction, verification, and comparison creates inspectable intermediate outputs. A single ranking prompt hides omissions and can overweight presentation quality.

4. A government benefits office wants to deploy Claude to help caseworkers draft eligibility determination letters. The agency has no dedicated AI engineering team and must avoid building custom middleware. Throughput is critical: caseworkers handle 200+ cases per day and cannot tolerate slow or inconsistent responses. Which prompt-engineering approach best balances performance and minimal-code constraints in this scenario?

A) Ask each caseworker to paste the full policy manual into every conversation so Claude always has complete context, relying on the long context window to compensate for the lack of a system prompt.

B) Build a retrieval-augmented pipeline with a vector database to inject only relevant policy clauses per case, minimizing token usage and maximizing throughput.

C) Create a Claude.ai Project with a detailed system prompt containing role definition, output format, and tone constraints, then save reusable instructions so every caseworker session inherits them without per-request setup.

D) Fine-tune a Claude model on historical determination letters to embed agency style, eliminating the need for explicit instructions in each session.

Answer: C. Projects let teams store persistent instructions and context that apply automatically to every conversation, removing per-request setup and requiring zero custom code - exactly matching the minimal-code constraint. Pasting the full policy manual each time is error-prone, slow, and inconsistent. A retrieval-augmented pipeline requires custom engineering infrastructure the agency lacks. Fine-tuning is not a self-service option available through Claude.ai and introduces significant engineering overhead.

DOMAIN 2 / OUTPUT EVALUATION AND VALIDATION

5. Claude drafts a market brief containing five precise growth figures. The brief will inform a budget decision. What should the analyst do before using it?

A) Regenerate the brief with a different tone

B) Trust the figures because they are precise

C) Ask Claude whether it is confident

D) Verify each figure against the cited primary source and mark unsupported claims

Answer: D. Precision is not evidence. High-impact factual claims should be checked against authoritative sources; self-reported model confidence is not a substitute for verification.

6. A team compares two prompt versions for customer-email drafts. Reviewers keep disagreeing about which is better. What is the best improvement to the evaluation?

A) Use whichever prompt produces longer emails

B) Define a rubric for factual accuracy, policy compliance, clarity, and actionability, then score a representative set blind

C) Let Claude choose its preferred prompt

D) Evaluate only the easiest customer request

Answer: B. A task-specific rubric and representative blind sample make judgments repeatable and expose tradeoffs. Length, model preference, and easy-only examples are weak proxies.

7. A health benefits team uses Claude to draft explanations of plan coverage. Which control is most appropriate before sending a response that could affect care decisions?

A) Require review by a qualified benefits specialist and link the answer to current plan documents

B) Use a more creative model setting

C) Ask the employee to decide whether the answer is accurate

D) Automatically send any answer that sounds certain

Answer: A. The consequence of an incorrect coverage answer warrants domain-expert review and grounding in current authoritative documents. Fluency or confidence does not lower the risk.

8. Claude produces a correct contract summary but fails to mention an unusual termination clause. How should the team evaluate future summaries?

A) Increase the output length

B) Accept the omission because the rest is correct

C) Check only whether the summary reads clearly

D) Use a checklist of required clause categories and compare each summary with the source passages

Answer: D. Completeness needs explicit coverage criteria and source comparison. A longer response may still miss the same high-impact clause.

9. A cybersecurity operations center uses Claude to generate real-time threat advisories from ingested log data. Analysts need sub-second response latency, making synchronous human fact-checking before delivery impractical. The team wants to design an evaluation framework that catches factual errors without adding perceptible delay. Which evaluation design best satisfies both constraints?

A) Run automated, offline evaluation pipelines that periodically score a sample of past advisories against authoritative threat-intelligence feeds, using findings to refine prompts and flag systematic error patterns.

B) Embed a second synchronous Claude call after each advisory to self-critique the output before delivery, doubling inference latency on every request.

C) Require a human analyst to approve every advisory before it reaches the dashboard, accepting that latency will exceed the sub-second target during high-volume incidents.

D) Disable factual evaluation entirely and rely on downstream SIEM correlation rules to catch inaccuracies after advisories are acted upon.

Answer: A. Offline, sampled evaluation pipelines decouple accuracy assessment from the live request path, preserving sub-second latency while still detecting systematic factual errors through comparison with authoritative sources - directly aligning with the agent evaluation principle of testing outputs against ground truth without blocking production. Requiring synchronous human approval violates the latency constraint. Disabling evaluation entirely removes the accuracy safeguard the domain requires. A second synchronous Claude call doubles latency and relies on the same model to self-correct, which does not constitute independent factual verification.

DOMAIN 3 / PRODUCT AND MODEL SELECTION

10. A policy team repeatedly asks Claude questions about the same approved handbook and wants shared instructions to persist across sessions. Which setup best fits?

A) Start an unrelated temporary chat for every question

B) Use a shared Project with the approved handbook and scoped project instructions

C) Paste the entire handbook into every prompt manually

D) Build a custom API integration before testing the workflow

Answer: B. A Project is designed for reusable knowledge and instructions around ongoing work. It reduces repeated setup without adding unnecessary application engineering.

11. A team has thousands of low-risk product descriptions to classify into a fixed taxonomy and a small number of difficult ambiguous cases. What is the best model strategy?

A) Route routine cases to a faster lower-cost model and escalate uncertain cases to a more capable model or human

B) Use the fastest model and never review results

C) Choose models based only on context-window size

D) Use the most capable model for every item regardless of cost

Answer: A. Routing by task difficulty preserves quality where it matters while controlling cost and latency. A single-model policy ignores the workload's clear risk tiers.

12. A product team is building a customer-facing support assistant using Claude. During peak hours, the assistant must respond quickly to simple FAQs, while complex billing disputes require thorough, high-accuracy answers. The team has a fixed monthly API budget and cannot afford to route every request through the most capable - and most expensive - model. Which approach best balances response quality, latency, and cost across both request types?

A) Cache the entire system prompt and all user messages so that every subsequent request is served from cache, eliminating per-request model inference costs entirely.

B) Use a routing workflow that classifies incoming requests and directs simple FAQ queries to a smaller, faster, lower-cost Claude model while escalating complex billing disputes to a more capable model.

C) Switch all requests to streaming mode so that the first token arrives faster, reducing perceived latency without changing the underlying model or cost structure.

D) Send every request to the most capable Claude model to guarantee maximum accuracy, and reduce costs by shortening all system prompts.

Answer: B. A routing workflow classifies each incoming request and directs it to the model best suited for that category. Simple, high-volume FAQ queries go to a smaller, faster, lower-cost model, preserving budget and minimizing latency for straightforward tasks. Complex billing disputes are escalated to a more capable model where accuracy justifies the higher cost. This directly balances quality, latency, and cost across distinct request types. Sending every request to the most capable model maximizes quality but ignores the cost ceiling entirely and does not address latency differences between request types. Prompt caching reduces redundant token processing costs for repeated context, but it does not eliminate per-request inference costs and cannot substitute for model selection decisions. Streaming affects how quickly the first token is delivered to the user but does not change which model processes the request, so it has no impact on cost or output quality.

13. Claude helps a sales team prepare account updates. It may summarize CRM notes but must not change opportunity stages without an account owner's consent. What design is best?

A) Allow draft recommendations, but require an explicit owner approval before any CRM write

B) Put the rule only in a friendly reminder

C) Disable CRM access and copy updates by hand forever

D) Let Claude update every stage and review changes monthly

Answer: A. The workflow separates low-risk analysis from a consequential write action and places approval at the action boundary. A prompt reminder alone is not enforcement.

14. A research workflow involves collecting sources, extracting claims, and writing a final memo. Where should human review add the most value?

A) Before any source is collected

B) Only when Claude reports low confidence

C) Only after the memo has been distributed

D) At a defined checkpoint that verifies source quality and claim support before synthesis

Answer: D. Reviewing evidence before synthesis prevents weak sources from contaminating the final memo. Model confidence is not a dependable trigger for all errors.

15. A team automates invoice intake with Claude. Most invoices are standard, but some lack purchase-order numbers. What is the safest efficient workflow?

A) Ask Claude to retry until it outputs a number

B) Invent a purchase-order number when one is missing

C) Process complete invoices automatically and route missing or conflicting fields to an exception queue

D) Reject every invoice

Answer: C. A deterministic exception path preserves automation for routine cases without fabricating required data. Retrying cannot recover information absent from the source.

16. An online marketplace uses a Claude-powered agent to auto-resolve buyer disputes. After three months, the team notices that roughly 12% of auto-resolved cases are later escalated by unhappy buyers. The engineering lead wants to improve the system but has no clear view of *why* those cases failed. Which evaluation approach best supports reproducible failure diagnosis while staying within the guidance for building effective agents?

A) Route all disputed cases to a human agent for 30 days to collect a clean baseline, then retrain the model before re-enabling automation.

B) Add a random 10% sampling audit where human reviewers re-read dispute transcripts and flag anything that 'feels wrong,' then count flags per week.

C) Increase the confidence threshold for automation so fewer cases are auto-resolved, reducing the absolute number of escalations without investigating root causes.

D) Define a structured scorecard with explicit criteria (e.g., policy rule applied, evidence cited, refund amount within bounds) so every auto-resolved case can be scored consistently and failing cases cluster into diagnosable categories.

Answer: D. Reproducible failure diagnosis requires explicit, consistent evaluation criteria - not subjective impressions or volume reduction. A structured scorecard maps each decision to verifiable criteria (policy rule, evidence, amount), enabling failures to cluster into actionable categories. Random subjective audits produce inconsistent signals. Raising the confidence threshold hides failures rather than diagnosing them. A 30-day human-only period is high-overhead and delays diagnosis. Anthropic's agent evaluation guidance emphasizes defining clear, measurable success criteria before diagnosing failures.

DOMAIN 5 / CONFIGURATION AND KNOWLEDGE MANAGEMENT

17. A shared Project contains three policy files with conflicting effective dates. Claude sometimes cites an obsolete rule. What should the owner do first?

A) Tell users to ignore wrong answers

B) Increase response length

C) Add more files so Claude sees every historical version

D) Remove or clearly archive obsolete material and label the current authoritative policy with its effective date

Answer: D. Knowledge quality starts with a clean source set. Explicit authority and dates reduce ambiguity better than asking the model to resolve undocumented conflicts.

18. A department has one instruction that applies to every task and another that applies only to press releases. How should they organize the guidance?

A) Rely on team members to remember both

B) Put both in every individual prompt

C) Keep the universal rule in shared project instructions and invoke the press-release guidance only for that workflow

D) Combine all possible rules into one very long instruction

Answer: C. Separating persistent universal guidance from task-specific instructions keeps context relevant and reduces conflicts.

19. A wealth-management firm uses Claude.ai on a Team plan. The knowledge base for their advisory work contains three documents: a compliance policy, a client-communication style guide, and a proprietary risk-scoring rubric. The compliance officer wants internal governance documents kept structurally separate from client-facing content so that advisors focused on client communications are never working in the same knowledge base as the internal risk-scoring logic. Which Claude Projects configuration approach most directly satisfies this structural content-separation requirement?

A) Store all three documents in a single Project knowledge base and rely on Claude's responses to indicate which document was referenced in each session.

B) Create two separate Projects - one whose knowledge base contains only the compliance policy and risk-scoring rubric for internal advisory work, and a second whose knowledge base contains only the style guide for client-communication drafting - so that each workspace holds only the documents relevant to its purpose.

C) Upload all three documents as file attachments inside individual conversations so that each conversation is self-contained and knowledge does not persist across sessions.

D) Place the compliance policy and risk-scoring rubric in the Project instructions field and put the style guide in the Project knowledge base, keeping all three documents within the same Project.

Answer: B. Claude.ai Projects are documented as self-contained workspaces, each with its own knowledge base and instructions. Creating two separate Projects - one holding only the internal compliance policy and risk-scoring rubric, and another holding only the client-facing style guide - establishes a hard structural boundary enforced by the Projects architecture itself. Each Project's knowledge base contains exclusively the documents relevant to that use case, so internal governance materials and client-facing content cannot intermingle within the same workspace. This is the configuration practice that most directly satisfies the structural separation requirement using native Projects design. Storing all three documents in a single mixed knowledge base eliminates any content boundary. Within a single Project there is no mechanism to restrict which uploaded documents are available in a given session, so all documents remain accessible to everyone in that workspace regardless of their role or focus. Uploading documents as per-conversation file attachments makes knowledge ephemeral and non-reusable across team members. This directly undermines the reusable, shared-workspace purpose that Projects are designed to serve, and it does not create a durable structural separation between internal and client-facing content. Placing some documents in the instructions field and others in the knowledge base still consolidates all materials into one Project. The instructions field is designed for behavioral guidance text, not for hosting full policy or rubric documents, and all content remains within the same workspace - so the structural separation requirement is not met.

20. An employee wants Claude to help categorize support cases. The raw export contains names, phone numbers, and account numbers that are unnecessary for categorization. What should happen first?

A) Delete the output after sharing it

B) Upload the full export because Claude needs realism

C) Remove or mask unnecessary personal data before providing the cases

D) Ask Claude not to mention the data

Answer: C. Data minimization reduces exposure at the source. A prompt instruction does not undo unnecessary disclosure of sensitive inputs.

21. A marketing group uses Claude to generate regulated product claims. Who remains accountable for the published claim?

A) Claude, because it generated the wording

B) The human and organization that approve and publish it

C) The customer who reads it

D) No one if the prompt included a disclaimer

Answer: B. People and organizations remain responsible for decisions and published content. AI assistance does not transfer legal or professional accountability.

22. A hiring-support prompt appears to rate candidates from one demographic lower. What is the best response?

A) Pause use, test a representative controlled set for disparate behavior, inspect inputs and criteria, and involve the responsible review team

B) Hide demographic fields only in the final report

C) Ask Claude to promise to be fair

D) Deploy it because the average rating is accurate

Answer: A. Potential discriminatory behavior requires a measured evaluation, root-cause review, and accountable oversight before continued use. A promise in the prompt is not evidence.

23. A subscription software company is building a Claude-powered support-ticket triage agent. The team wants to optimize response quality over time, but their evaluation results are inconsistent across sprint cycles because different engineers run ad-hoc tests against different ticket samples. Which single practice most directly establishes a reproducible evaluation baseline that the team can track across releases?

A) Define a fixed, versioned golden dataset of representative support tickets with expected triage labels, and run the same evaluation harness against every new model or prompt version.

B) Increase the number of tickets sampled per evaluation run each sprint so that statistical variance decreases naturally over time.

C) Switch from human-reviewed labels to automated keyword matching so that evaluation scores are computed deterministically without human subjectivity.

D) Rotate the engineers who run evaluations each sprint to reduce individual bias and ensure diverse perspectives on quality.

Answer: A. A fixed, versioned golden dataset with a consistent harness is the foundation of reproducible evaluation; without it, each run measures a different distribution and results cannot be compared across releases. Increasing sample size each sprint does not fix the core problem - varying samples still prevent apples-to-apples comparison. Rotating engineers addresses reviewer bias but not dataset inconsistency. Automated keyword matching trades accuracy for determinism and cannot capture nuanced triage quality, undermining the goal of performance optimization.

DOMAIN 7 / TROUBLESHOOTING AND OPTIMIZATION

24. Claude returns inconsistent meeting-note action items across similar transcripts. What is the most useful first diagnostic?

A) Change several settings at once

B) Collect failing examples and compare them with passing ones against an explicit extraction rubric

C) Assume the model is random and stop testing

D) Make the prompt longer without examining failures

Answer: B. A small failure set and rubric reveal whether the issue is missing instructions, ambiguous source text, or evaluation drift. Changing many variables hides causality.

25. A Project response has become slower and less focused after months of adding background files. What is the best first optimization?

A) Audit the knowledge set, remove redundant or stale material, and keep only task-relevant authoritative sources

B) Ask for twice as much output

C) Create more conflicting instructions

D) Add the same files again

Answer: A. Relevant, curated context is usually more useful than maximum context. Redundant and stale material increases distraction and ambiguity.

Want the rest?

The full CCAO-F practice test on learn.cloudyeti.io has 150 questions, a timed exam mode, and a score by domain so you know what to study next.

Start the free CCAO-F practice test

These are original practice questions written by CloudYeti. They are NOT real exam questions and are not affiliated with or endorsed by Anthropic.