

Boost up Your Certification Score

Huawei H13-323

Huawei Certified ICT Professional-AI Solution Architect



For More Information – Visit link below:

<https://www.examsboost.com/>

Product Version

- ✓ Up to Date products, reliable and verified.
- ✓ Questions and Answers in PDF Format.

Visit us at: <https://www.examsboost.com/test/h13-323>

Latest Version: 6.0

Question: 1

A financial institution is designing an assistant that drafts client communications. Drafts must reflect current product terms, must never be sent without an adviser's approval, and every sent communication must be reconstructable at audit two years later.

Which design satisfies all three requirements?

- A. Fine-tune the model on past approved communications and the current product terms, require adviser approval before any draft is sent, and retain a record of the model version used to produce each individual communication
- B. Ground drafts in the current product terms through retrieval, allow advisers to send an approved draft directly from within the assistant, and retain the full conversation history for later audit purposes if a communication is queried
- C. Ground drafts in the current terms through retrieval, route every draft through an explicit adviser approval step, and record the request, the retrieved sources, the model version and the approved text for each communication
- D. Ground drafts in the current product terms through retrieval, require an explicit adviser approval before sending, and rely on the model provider's own request logs to reconstruct communications at audit two years afterwards

Answer: C

Question: 2

An architect is planning the parallelism strategy for a large training job on a cluster of multi-accelerator nodes.

Where should tensor parallelism be applied?

- A. Within a node, across the accelerators that share the high-bandwidth internal interconnect
- B. Across nodes, so that the largest possible group of accelerators can share each layer's computation
- C. Across racks, so that layer computation is distributed evenly over the whole cluster
- D. Wherever the scheduler places the job, since the framework abstracts the physical topology from the parallelism configuration

Answer: A

Question: 3

A vendor claims its model is "state of the art across all major benchmarks" and offers a demonstration on its own curated examples.

What should the architect ask for that would most improve the basis for a decision?

- A. The full list of benchmarks on which the claim rests, together with the reported scores
- B. The parameter count and training corpus size, so capability can be compared against alternatives
- C. A reference customer in the same industry who will describe their deployment experience
- D. A run against a held-out sample of the organization's own data, scored on the metric the business case names

Answer: D

Question: 4

A hospital group's AI steering committee wants a single metric to report to its board each quarter for a deployed clinical documentation assistant.

Which measure most directly reflects delivered value?

- A. The total number of documentation requests the assistant serves each day across the hospital group
- B. Clinician minutes saved per documented encounter, measured against a pre-deployment baseline
- C. The proportion of the generated documentation text that clinicians accept without any editing
- D. The average latency between a clinician's request and the assistant's first returned token, measured across the group

Answer: B

Question: 5

An organization's assistant is grounded on a corpus that includes both current and superseded versions of its operating procedures. Users report confidently stated answers that reflect withdrawn procedures.

The team proposes instructing the model to prefer recent documents.

How should the architect respond?

- A. The corpus should be curated so that superseded procedures are excluded or clearly marked, because a retrieval and generation step cannot reliably enforce a currency rule the data does not express
- B. The instruction should be adopted as proposed, because the model is able to use the document dates or version labels present in the retrieved passages to identify and select the current procedure
- C. The retrieval step should be configured to return more passages, so the current procedure is likely to appear alongside the superseded one and the model can compare the two directly before answering
- D. A disclaimer should be added to every answer that the assistant returns, advising users to verify the procedure against the current handbook, so that the residual risk sits with the reader and not with the organization that deployed the assistant, which cannot check every answer

Answer: A

Question: 6

A retrieval-grounded assistant returns answers that are well written but frequently address a different aspect of the question than the user asked about. Inspection shows the retrieved passages are on the right general topic but do not contain the specific information requested.

Where does the fault lie?

- A. In the request format, since the user's question is not being passed across in a form that the retrieval step is able to match against
- B. In the generation step, since the model ought to recognize that the passages it was given do not answer the question and say so to the user directly
- C. In the corpus, since it evidently does not contain the specific information that the assistant's users are actually asking their questions about
- D. In the retrieval step, since the model can only answer from what it is given and topically similar passages are not the same as relevant ones

Answer: D

Question: 7

A company must decide between building its own training capacity and renting accelerator capacity for a program with one large adaptation run planned this year and no confirmed pipeline beyond it.

Which recommendation is best supported?

- A. Build capacity now, because owning the estate removes the data residency questions that renting introduces
- B. Build capacity now, because owned accelerators cost less per unit of compute than rented capacity once the estate is in place
- C. Rent capacity for this run, because a single known workload with no confirmed pipeline does not justify committing to an estate that must be utilized continuously to be economic
- D. Defer the run until a pipeline of further workloads has been confirmed, so that the capacity decision can be made once against the full demand profile rather than revisited later

Answer: C

Question: 8

An adapted model performs far better on the team's evaluation set than a subsequent review of live output suggests it should.

What should be checked first about the data?

- A. Whether the evaluation set is large enough to produce a stable estimate, since a small set can report an optimistic score by chance
- B. Whether examples from the evaluation set also appear in the adaptation data, since overlap makes the evaluation a measure of recall rather than capability
- C. Whether the evaluation set was labeled by the same people who wrote the adaptation data, since shared assumptions can make the scoring generous
- D. Whether the evaluation set reflects the current live distribution, since traffic may have shifted since the set was assembled

Answer: B

Question: 9

Two candidate models are being assessed for a summarization workload. Model P scores higher on an automatic overlap metric against reference summaries. Model Q scores lower on that metric, but reviewers consistently prefer its output, and its summaries omit details the references included while adding correct details the references omitted.

How should the architect interpret this evidence?

- A. The reviewers are applying inconsistent criteria to the two candidates, so their stated preference should be discounted until a formal agreement study establishes that they are all scoring the same property
- B. Model P should be selected on the strength of its metric score, because an automatic measure is reproducible on demand and defensible in review, while a reviewer preference cannot be audited or repeated at scale across a large evaluation set
- C. The overlap metric rewards resemblance to the reference text, so it penalizes a summary that is different and better; the reviewer preference is the stronger evidence and the metric should be supplemented
- D. The reference summaries are evidently of poor quality, so both of the reported scores are meaningless and a fresh reference set must be authored before either candidate model can be chosen for the workload

Answer: C

Question: 10

An energy utility is documenting the practical significance of multimodal capability for its inspection workflow.

Which statement is accurate?

- A. Accuracy on each individual input type improves relative to a specialist system, because training across several modalities gives the model a larger body of data, and a more varied one, to learn from

- B. A single model can reason over an inspection photograph and its accompanying written report together, removing the need to stitch separate vision and language systems into a pipeline
- C. Inference cost falls, because one model removes the need for the several specialist models a pipeline would otherwise require
- D. Images can be used without preprocessing, because the model accepts raw sensor output in whatever form the inspection device produces

Answer: B

Thank You for Trying Our Product

For More Information – **Visit link below:**

<https://www.examsboost.com/>

15 USD Discount Coupon Code:

G74JA8UF

FEATURES

- ✓ **90 Days Free Updates**
- ✓ **Money Back Pass Guarantee**
- ✓ **Instant Download or Email Attachment**
- ✓ **24/7 Live Chat Support**
- ✓ **PDF file could be used at any Platform**
- ✓ **50,000 Happy Customer**



Visit us at: <https://www.examsboost.com/test/h13-323>