CERTLABZ CERTIFICATION BOARD

Certified AI Risk and Ethics Professional (CAREP)

Ten Sample Questions

These ten questions are written at the difficulty of the live exam. Six carry a figure, a data table, a configuration or a code listing; four are plain scenario questions, as on the live paper. Every one is a scenario a working AI governance professional would meet, and none of them appears in the live question bank.

The live exam is multistage adaptive: how you score on one stage selects the difficulty of the next, so no two candidates sit the same paper. It is a performance-based exam: each form can carry up to 5 to 12 performance-based questions alongside the multiple-choice questions, which means you build, sequence or complete something rather than only choosing from a list. Answers and full explanations follow each question here; the live exam does not disclose item-level answers.

Question 1 — Domain 1: AI Risk and Ethics Foundations — DATA

Emma Wilson maps responsible-AI principles to what they require.

Principle	Requirement
Fairness	Outcomes do not unjustly disadvantage groups
Transparency	People can understand and contest decisions
Accountability	Clear ownership and answerability
Safety	The system avoids and mitigates harm

Table 1.1 Principles and requirements.

A model is highly accurate overall but far worse for one protected group. Which principle is most at risk, and why?

A. Safety, because a model that is accurate overall cannot cause any harm and the subgroup difference is not a real concern at all.

B. Fairness, because high aggregate accuracy can hide serious unfairness to a subgroup, which ethics and often law require you to address.

C. Transparency, because the only issue with a subgroup difference is that it has not yet been disclosed to the people affected by it.

(D) Accountability, because as long as someone owns the model its worse performance for one group is automatically acceptable and compliant.

Correct answer: B
High overall accuracy can mask serious harm to a subgroup, which is a fairness problem that ethics and often anti-discrimination law require you to investigate and address. It is not made acceptable by good aggregate numbers, by disclosure alone, or simply by having an owner.

Question 2

Domain 2: Fairness, Bias and Harm

CONFIG

James Carter reviews a lending model built to be fair.

```
training data: past approvals, historically skewed
mitigation   : removed the "race" field
result       : approval rates still differ sharply by race
features     : include postcode, which correlates with race
```

Listing 2.1 The model and its data.

Why does bias persist, and what is the right response?

(A) Removing the race field made the model fair, so the remaining difference must be a coincidence that requires no further action at all.

(B) Proxy features such as postcode reintroduce race, so bias persists; test outcomes across groups and mitigate rather than just drop the field.

(C) The model needs more historical approval data, because training on even more of the past skewed decisions will make the outcomes fair.

(D) The difference is acceptable because the race field was removed, and any disparity that remains is therefore not the model's responsibility.

Correct answer: B
Dropping the race field does not remove bias because proxies such as postcode still carry race, and training on skewed history entrenches past discrimination. The right response is to test outcomes across groups and actively mitigate, not to assume fairness because a field was removed or to add more biased history.

Question 3

Domain 3: Transparency, Explainability and Accountability

DATA

Olivia Davis reviews an automated decision system.

Aspect	Current state
Decision	Automated refusal of a service
Explanation to the person	None given
Appeal route	None available
Record of data and logic used	Not retained

Table 3.1 The system as deployed.

Which principles are violated, and what is needed?

A. None are violated, because a fully automated refusal with no explanation or appeal is an efficient and acceptable design for any decision.

B. Transparency and accountability are violated; the person needs an explanation and an appeal route, and the decision must be traceable.

C. Only fairness is violated, and the fix is simply to make the model more accurate overall so that fewer people are ever refused service.

D. Only safety is violated, and the fix is to speed up the automated refusal so that affected people receive the outcome more quickly.

Correct answer: B
A consequential automated refusal with no explanation, no appeal and no retained record breaches transparency and accountability, and often legal rights. People must be able to understand and contest the decision, and the organisation must be able to trace the data and logic behind it. Accuracy and speed do not address these gaps.

Question 4

Domain 4: AI Governance and Regulation

CONFIG

Liam Anderson lists the steps to stand up AI governance, shown out of order.

```
A. Apply proportionate policies and oversight
B. Inventory the AI systems in use
C. Monitor, review and respond to incidents
D. Assess and classify each system by risk
```

Listing 4.1 Steps to order.

What is the correct order?

A. A, then C, then B, then D, applying oversight and monitoring before anyone has catalogued or risk-assessed the AI systems in use.

B. B, then D, then A, then C, inventorying systems, classifying them by risk, applying proportionate oversight, then monitoring and responding.

C. C, then A, then D, then B, monitoring for incidents before the systems have been inventoried, classified or given any oversight at all.

D. D, then B, then C, then A, classifying systems before they have been inventoried and monitoring before any oversight is in place.

Correct answer: B
Governance starts with an inventory of AI systems (B), then risk classification (D), then proportionate policies and oversight (A), then ongoing monitoring and incident response (C). You cannot assess, control or monitor systems you have not first catalogued, so the other orders put later steps before their prerequisites.

Question 5

Domain 5: Risk Management Across the AI Lifecycle

DATA

Ava Thompson maps lifecycle risks to controls.

Risk	Control
Accuracy for a group declines over a year	(to choose)
A data update causes harmful outputs	(to choose)
Unusual or malicious inputs	(to choose)
A high-impact wrong decision	(to choose)

Table 5.1 Lifecycle risks.

Which set of controls fits these risks?

A. A one-time launch test, deploying updates untested, testing only typical inputs, and full automation with no human review of any decision.

B. Ongoing disaggregated monitoring, testing updates with a rollback plan, adversarial and edge-case testing, and human override of decisions.

C. No monitoring, deploying updates immediately, skipping input testing, and removing human oversight so decisions are made faster overall.

D. Checking overall accuracy only, trusting all updates, testing only common inputs, and letting the model approve its own high-impact decisions.

Correct answer: B
Silent per-group decline needs ongoing disaggregated monitoring, a risky update needs pre-release testing plus a rollback plan, unusual or malicious inputs need adversarial and edge-case testing, and high-impact decisions need human override. The other options describe exactly the deploy-and-forget failures that cause AI harm.

Question 6

Domain 6: Privacy, Security and Responsible Deployment

CONFIG

A team plans to train a model on customer records and deploy it publicly.

```
training data : personal customer records
lawful basis  : not established
notice        : customers not informed
risk          : model may reveal training data when prompted
```

Listing 6.1 The plan.

What must the team address before proceeding?

A. Nothing, because training a model on personal data is exempt from privacy law and the leakage risk is only a theoretical concern.

B. Establish a lawful basis and inform customers, and mitigate the data-leakage risk, since it is both a privacy harm and a security vulnerability.

C. Only the model accuracy, because once the model performs well the missing lawful basis and the data-leakage risk both cease to matter.

D. Only the deployment speed, because getting the model live quickly is what reduces the chance that the training data is ever leaked at all.

Correct answer: B
Training on personal data needs a lawful basis and transparency to customers, and a model that can reveal its training data is both a privacy harm and a security vulnerability that must be mitigated. These are prerequisites, not optional extras, and neither accuracy nor speed removes them.

Question 7

Domain 1: AI Risk and Ethics Foundations

A team says "our model is just a tool, so any harm from how it is used is entirely the user's responsibility". Why is this inadequate?

A. It is correct, because the people who design and deploy an AI system bear no responsibility for foreseeable harms once it is in a user's hands.

B. Designers and deployers share responsibility for foreseeable harms and misuse, and for building in safeguards, so they cannot disclaim it entirely.

C. It is correct, but only for low-risk systems, because responsibility for harm transfers to the user the moment any tool is made available to them.

D. It is inadequate only because the statement was not written down in a policy, and documenting the same position would make it fully acceptable.

Correct answer: B
Responsible AI holds designers and deployers accountable for foreseeable harms and misuse and for building in safeguards; they cannot simply disclaim all responsibility by calling the system a tool. This applies regardless of risk level, and writing the disclaimer into a policy would not change the underlying accountability.

Question 8

Domain 2: Fairness, Bias and Harm

A hiring model is trained on a company's past hires, who were overwhelmingly from one demographic. What is the main risk?

A. There is no real risk, because training on the company's own historical hiring decisions guarantees that the model will be fair to everyone.

B. The model may learn and perpetuate the historical bias, disadvantaging under-represented candidates, so it must be tested and mitigated.

C. The only risk is that the model will be slightly less accurate, which can be fixed simply by training it on even more of the same past hiring data.

D. The risk is purely technical performance, because a model trained on real hiring decisions cannot by its nature produce any unfair outcomes.

Correct answer: B
A model trained on skewed historical hiring decisions is likely to learn and reproduce that bias, disadvantaging under-represented candidates, which is a classic way AI entrenches discrimination. It must be tested across groups and mitigated. Training on more of the same biased history makes it worse, not fair.

Question 9 Domain 4: AI Governance and Regulation

An organisation buys an AI model from a vendor and uses it to make regulated decisions about customers. Who is accountable for compliance?

A. Only the vendor, because the organisation simply purchased the model and therefore carries no responsibility for how it is applied to customers.

B. The deploying organisation remains responsible for how it uses the system and its outcomes, alongside the vendor's own obligations.

C. No one is accountable, because responsibility for an AI decision cannot be assigned once a third-party model is involved in making it at all.

D. Only the individual customer, because by using the service they accept full responsibility for whatever decision the AI model produces about them.

Correct answer: B
Buying a model does not outsource accountability: the deploying organisation is responsible for how it uses the system and for the outcomes, alongside the vendor's obligations. Responsibility does not vanish because a third party built the model, and it certainly does not fall on the customer.

Question 10 Domain 5: Risk Management Across the AI Lifecycle

A model performed well in testing but fails badly on a real population it was never evaluated on. What lifecycle gap does this reveal?

A. No gap exists, because strong results in pre-launch testing always guarantee that a model will perform well on every real population it meets.

B. Evaluation did not reflect the real deployment population; testing must match actual use, and production monitoring must catch what testing missed.

C. The only gap is model size, and using a larger model would have guaranteed good performance on the new population without any other change.

D. The gap is purely that the model was launched too slowly, and a faster launch would have prevented the failure on the real population entirely.

Correct answer: B
A model can pass testing yet fail on a population it was never evaluated against, which shows the evaluation did not reflect real deployment conditions. Testing must match actual use, and production monitoring must catch failures that pre-launch testing missed. Model size and launch speed do not address an unrepresentative evaluation.

Published by the CertLabz Certification Board, a division of CertLabz LLC. Companion to objectives document CAREP-001. https://certlabz.com/certifications/certified-ai-risk-and-ethics-professional.html