

PRACTICE QUESTIONS

AI CERTs

# AI+ Ethics

SHARED BY [u/Silent-Badger-9528](#) on aiexamcerts

|           |                     |
|-----------|---------------------|
| EDITION   | Free Set 1          |
| QUESTIONS | 10                  |
| SHARED    | October 15, 2026    |
| EXAM      | AI CERTs AI+ Ethics |

---

Shared by [u/Silent-Badger-9528](#) on aiexamcerts

10 practice questions for the AI CERTs AI+ Ethics exam, each followed by its answer and a short explanation. More free sets and the full practice exam: [www.aiexamcerts.com/ai-certs/ai-ethics](https://www.aiexamcerts.com/ai-certs/ai-ethics). Spotted a mistake? Email [support@aiexamcerts.com](mailto:support@aiexamcerts.com) with the exam name and the question number.

---

1. Which term describes an AI system's ability to keep performing reliably when inputs differ from its training data or are deliberately perturbed?
- A Interpretability of learned weights
  - B Robustness
  - C Horizontal scalability
  - D Inference cost optimization

**Answer: B**

Robustness is the property that a system keeps reliable behavior under distribution shift, noisy inputs, or adversarial perturbation. Interpretability concerns understanding internal logic, a different property even though it supports debugging. Horizontal scalability describes adding capacity, not stability when inputs change.

---

2. The practice of having a dedicated group try to break a system with adversarial inputs before release is called red \_\_\_\_\_.

**Answer: teaming**

Red teaming uses an adversarial group to probe a system for harmful outputs, misuse paths, and security weaknesses before it ships. The exercise complements standard testing because it looks for failures a cooperative test set would miss.

---

3. A team ships a chatbot without documenting its intended use or known limitations. Which lifecycle practice did it most clearly skip?
- A Hyperparameter tuning sweeps
  - B Load balancing across regions
  - C Dataset deduplication before training
  - D Documentation of intended use and limits

**Answer: D**

Documenting intended use, out-of-scope uses, and known limitations is a core responsible AI practice that helps deployers and users avoid misuse. Hyperparameter tuning and load balancing are model and infrastructure steps that do not communicate boundaries. Dataset deduplication is the most tempting wrong answer because it is a data quality step, but it does not tell users where the system should not be used.

---

4. Documenting a model's purpose, data sources, and known limitations is a recognized transparency practice.
- A True
  - B False

**Answer: A** (True)

Model documentation such as model cards and datasheets lets deployers and auditors judge whether a system fits their context. Responsible AI frameworks widely recommend recording purpose and limits. The statement is true because disclosing limitations supports informed use, even though documentation alone does not prove the model is accurate.

5. Which two practices make an automated decision easier to explain to the person it affects? (Choose two.)
- A Hiding the model type from reviewers
  - B Replacing reasons with a generic rejection code
  - C Logging the reason codes behind each adverse decision
  - D Increasing the model's parameter count
  - E Providing a plain-language summary of the main factors

**Answer: C, E**

Recording reason codes lets staff reconstruct why a specific adverse decision happened, and a plain-language summary gives the affected person something they can act on. Hiding the model type and replacing reasons with a generic code reduce explainability. Parameter count is unrelated to explaining a decision, and the generic rejection code is tempting only because it looks tidy.

6. Which principle requires that a person affected by an automated decision can challenge it and obtain a meaningful review?
- A Contestability and recourse
  - B Model compression for mobile devices
  - C Data residency for cloud storage
  - D Differential privacy for query results

**Answer: A**

Contestability and recourse give affected people a route to challenge an automated decision and have it reviewed, often by a human. Model compression is a technical optimization, and differential privacy protects data rather than giving people a route to challenge a decision. Data residency is the closest distractor because it concerns legal control of data, but it does not provide a way to challenge a decision.

7. A dataset has few records for one demographic group. What is the most direct first step to address the representation gap?

- A Report only the overall accuracy for the dataset
- B Remove the group's field from the logs
- C Collect more representative data
- D Reduce the number of input features

**Answer: C**

Representation gaps are addressed at the source by collecting, sourcing, or carefully augmenting data for the under-represented group. Reporting only overall accuracy or removing log fields hides the problem rather than fixing it. Removing the group's field looks like neutrality, but it does not add the missing records.

8. Under the EU AI Act, systems that score individuals' social behavior to produce detrimental treatment are classified as \_\_\_\_\_ risk and prohibited.

**Answer: unacceptable** (also accepted: unacceptable risk)

The EU AI Act places social scoring that leads to detrimental or unjustified treatment in its unacceptable risk category, which bans those practices. Lower tiers such as limited and minimal risk carry lighter or no obligations. High-risk is the tempting answer because high-risk systems are allowed under heavy regulation, while this category is prohibited outright.

9. Which step most directly reduces the chance that a feature acts as a proxy for a protected attribute such as race?

- A Adding more hidden layers to the network
- B Enabling gradient clipping during training
- C Lowering the batch size
- D Testing features for proxy correlation

**Answer: D**

Checking whether a feature correlates with a protected attribute, and removing or constraining it when it does, addresses proxy discrimination at the input. Hidden layers, gradient clipping, and batch size tune optimization but do not change which signals the model can use. Adding hidden layers is tempting because architecture changes feel like fixes, but they make bias harder to inspect.

10. Which term describes an explanation that is accurate for one particular prediction but says nothing about how the whole model behaves?
- A Local explanation
  - B Global surrogate
  - C Model fingerprint
  - D Gradient checkpoint

**Answer: A**

A local explanation describes why the model produced one particular output, such as attributing that decision to specific input features. A global surrogate approximates the whole model, which is the opposite scope. Global surrogate is the most tempting wrong answer because it is also an explanation method, but its scope is the entire model.