1 0.4 / 0.4 points

Which of the following emerging AI concerns was explicitly discussed in Lecture 5 as part of ethical AI governance?

Hint: Think about GDPR, CCPA, and similar privacy regulations.

- Reducing the power consumption of GPUs in AI training
- Quantum computing's impact on AI models
- ✓ Privacy breaches due to generative AI
- AI's role in cryptographic security

2 0.4 / 0.4 points

Which of the following best describes the primary goal of AI in fairness metrics?

- To increase the computational efficiency of models
- To ensure AI models achieve higher accuracy
- ✓ To mitigate bias across different demographic groups
- To prevent adversarial attacks in LLMs

3 0.4 / 0.4 points

Which AI governance framework was mentioned in Lecture 5 as an example of corporate AI ethics initiatives?

- EU AI Act Compliance Framework
- OpenAI's Alignment Committee
- Google's "AI First" initiative
- ✓ IBM's AI Ethics Board

4 0.4 / 0.4 points

In a fairness-aware AI pipeline, which of the following is an example of an in-processing technique?

Hint: In-processing means modifying the model during training.

- Removing biased samples before training
- Post-processing to adjust model decisions
- Re-weighting input training data
- ✓ Adversarial Debiasing during model training

5 0.4 / 0.4 points

Given a Transformer-based model, which of the following modifications would best reduce bias in its out puts?

Hint: Think about fairness-aware ML strategies.

- Using dropout layers to prevent overfitting
- ✓ Applying re-weighting strategies to training samples
- Increasing the number of training epochs
- Using a larger dataset without filtering

6 0.4 / 0.4 points

Consider the following pseudo-code for a simple bias aware decision-making system:

```
if sensitive_feature in input_data:
 adjust_weight(input_data)
 decision = model.predict(adjusted_input)
else:
 decision = model.predict(input_data)
```

Which fairness strategy is being applied here?

Hint: Adjusting input features for fairness relates to this method.

- Adversarial Debiasing
- Equalized Odds Enforcement
- ✓ Disparate Impact Mitigation
- Counterfactual Fairness

7 0.4 / 0.4 points

Large language models (LLMs) like BERT are prone to which of the following ethical risks?

Hint: Ethical concerns go beyond just bias.

- Inability to perform fine-tuning on new tasks
- ✓ Hallucination and generating misleading information
- Reduction in computational efficiency
- Automatic reduction of bias due to large-scale training

8 0.4 / 0.4 points

The Belmont Report outlines three ethical principles for AI and research involving humans. Which of the following is NOT one of them?

Hint: One of these is a broad AI principle rather than a Belmont Report principle.

- Autonomy
- Beneficence
- Justice
- ✓ Fairness

9 0.4 / 0.4 points

The AI Fairness 360 (AIF360) toolkit provides various bias mitigation techniques. Which of the following is NOT a method mentioned in Lecture 5?

- Adversarial Debiasing
- ✓ Explainability through SHAP
- Disparate Impact Removal
- Equalized Odds Post-processing

10 0.4 / 0.4 points

Consider the following Python snippet related to bias mitigation. What does it likely implement?

```
from aif360.algorithms.preprocessing import Reweighing
rw = Reweighing()
transformed_data = rw.fit_transform(dataset)
```

Hint: The method comes from AIF360's set of fairness algorithms.

- A way to reduce overfitting in neural networks
- A reinforcement learning approach to bias correction
- A technique to improve the generalization of a model
- ✓ A method to balance class distribution in AI fairness