1   1 / 1 point

Why has "Responsible AI" become a critical area of focus in recent years?

- ✓ ◉ Because algorithms are increasingly used in high-stakes scenarios that directly influence people's lives.

- ○ Because AI models have become too complex for human oversight.

- ○ Because regulatory bodies have mandated strict ethical guidelines for all AI development.

- ○ Because computational power has only recently allowed for fairness checks.

2   1 / 1 point

In the healthcare example discussed, why did the algorithm for identifying patients for managed care inadvertently disadvantage Black patients?

- ○ The dataset lacked sufficient representation of Black patients, leading to poor predictions for this group.

- ✓ ◉ The algorithm used healthcare spending as a proxy for sickness, and Black patients, despite being equally sick, had historically lower healthcare spending due to access issues.

- ○ Black patients were inherently less likely to benefit from managed care.

- ○ The algorithm was explicitly coded to consider race as a factor.

3 1 / 1 point

The example of the video interview scoring application, where wearing glasses or a scarf altered personality scores, was used primarily to illustrate:

○ The necessity of diverse training data for facial recognition software.

○ The advanced capabilities of AI in nuanced human assessment.

○ The importance of high-quality video for accurate AI analysis.

✓ ⦿ The need for users to be skeptical of AI applications, especially if they seem too good to be true.

4 1 / 1 point

What does the phrase "human in the loop" refer to in the context of responsible AI?

○ The process of humans manually labeling all training data for AI models.

○ Allowing users to provide direct feedback to the AI system to improve its performance over time.

✓ ⦿ Combining human judgment and oversight with the quantitative outputs of AI models to make final decisions.

○ The ethical requirement for human programmers to be accountable for AI-driven errors.

The module notes that "today's decisions become tomorrow's training data." What is the primary concern associated with this feedback loop in the context of algorithmic bias?

○ It reduces the model's ability to generalize to new, unseen populations.

○ It makes models overly sensitive to recent trends.

○ It increases the computational cost of retraining models.

✓ ● It can amplify existing biases, as biased decisions fed back into the system will reinforce and potentially worsen the bias in future iterations.

When discussing strategies to combat bias, what is the first step recommended before even starting to run models?

○ Implementing robust real-time model monitoring systems.

○ Immediately collecting as much diverse data as possible.

○ Selecting the most advanced and fair algorithm available.

✓ ● Pausing to review the problem scope, considering who will be affected and if the right problem is being solved inclusively.

7 1 / 1 point

If you were considering adopting an AI model from a vendor, which question would be most pertinent to ask regarding its applicability to your specific population?

○ "What programming language was used to develop this model?"

○ "How many parameters does your model have?"

✓ ○ "Will the model be retrained to fit my organization's population, and how will you monitor its fairness for my specific groups?"

○ "Can you provide a discount if we commit to a long-term contract?"

8 1 / 1 point

What is the core idea behind "individual fairness" as a concept in AI?

○ Ensuring that the AI model performs with equal accuracy across all demographic groups.

○ Allowing each individual to customize the AI's decision-making process for themselves.

○ Guaranteeing that every individual affected by an AI decision has the right to an explanation.

✓ ○ Treating similar individuals similarly, irrespective of their demographic characteristics, when they possess comparable relevant attributes.

9 1 / 1 point

"Group fairness" in AI is primarily concerned with:

○ The subjective feeling of fairness experienced by the largest group affected by the AI.

○ The model provides identical outcomes for all individuals within a specific protected group.

✓ ● How the model's decisions and performance metrics (e.g., error rates like false positives) compare across different demographic groups.

○ Ensuring that the development team for the AI is demographically diverse.

10 1 / 1 point

What was the key finding from the experiment where ChatGPT was asked to complete sentences starting with "why doesn't he..." versus "why doesn't she..."?

○ The completions were nearly identical, showing the LLM's neutrality.

○ The LLM provided more creative and complex completions for the female pronoun.

✓ ● The autocompletions were significantly different, reflecting underlying patterns and potential biases in the training data concerning gender.

○ The LLM refused to complete sentences that could perpetuate stereotypes.