Teacher Answer Key

Case Study

1. The AI's training data came primarily from the internet, where Tigrinya-language content is extremely limited compared to English and other major languages. Because the model had seen very few examples of Tigrinya text, traditions, and cultural content during training, it did not have enough patterns to give accurate or detailed responses about that language and culture.

2. Amara's experience shows that AI tools reflect the biases of their training data, performing well on topics that dominate the internet (mainly Western, English-language content) and poorly on underrepresented languages, cultures, and histories. Users should not assume AI is equally reliable across all topics and should always seek specialized, community-based sources for cultural and linguistic research.

Multiple Choice

1. C
2. B
3. D

Matching

1-B, 2-A, 3-E, 4-C, 5-D

Fill-in-the-Blank

1) bias, 2) lawsuit, 3) code

Short Answer (Sample Responses)

Q1: Raw Common Crawl data contains spam, duplicate pages, auto-generated garbage, and harmful content, making it unsuitable for direct training. AI companies apply filters to remove low-quality content, duplicates, and sites associated with harmful material, then may further curate the data to improve balance and reduce bias.

Q2: Code follows strict formal rules, patterns, and logic structures, and training on it exposes the model to billions of examples of structured reasoning. These patterns appear to reinforce the model's ability to handle structured thinking in other domains, which is why models trained on code tend to perform better on math and logical reasoning tasks than those trained on text alone.

Q3: Answers will vary. Students might suggest actively seeking out and including more non-English web content, partnering with communities to create training data in underrepresented languages, weighting diverse sources more heavily, and building

separate evaluation benchmarks for different language groups to ensure quality across all populations the AI serves.