Teacher Answer Key

Case Study

1. An edge case is a rare or unusual scenario that doesn't appear often enough in training data for the AI to have learned how to handle it confidently. The mattress is a perfect edge case — it's not a pedestrian, not a vehicle, not a standard road obstacle the AI was trained on extensively. The system could classify it as 'unknown' but couldn't make a confident decision about how to respond, which is exactly the dangerous gap the reading describes.

2. 'For now' suggests the company views human monitors as a temporary bridge — present while the AI develops, but ultimately expected to be phased out. Students may reasonably argue this is responsible if the company is honest about the AI's limitations, or irresponsible if the timeline for removing humans is driven by cost reduction rather than genuine safety readiness. The Cruise incident (pedestrian dragged) is a relevant counterexample to reference.

Multiple Choice

1. A
2. D
3. C

Matching

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

Fill-in-the-Blank

1) 2, 2) defined, 3) Traffic

Short Answer (Sample Responses)

Q1: Tesla bets that cameras plus machine learning trained on billions of real-world miles can generalize to any road condition — a scalable but riskier approach. Waymo bets on LiDAR plus highly detailed maps of specific zones — more reliable in those zones but harder to scale geographically. Tesla's approach could eventually work everywhere; Waymo's requires enormous mapping investment for each new location.

Q2: When a collision is unavoidable, the AI must choose between protecting the passenger and protecting a pedestrian — a moral dilemma with no clear algorithmic answer. It's difficult because different ethical frameworks (utilitarian, rights-based) produce different answers, and because encoding any answer into a car means

making a societal-level moral decision by corporate fiat. Governments, ethicists, and the public should all have input, but no formal process for that exists yet.

Q3: The data is genuinely mixed. Waymo's crash data and Tesla's metrics show improvement over human averages in specific conditions, which is encouraging. However, the fact that both operate in favorable conditions (highways, geofenced zones) makes direct comparison complicated, and high-profile failures like the Cruise incident reveal that AI systems can fail in ways that would be preventable by human judgment.