Type 1. Perceptual Error

Case: Car Crash

Video LLaVA: The video shows a traffic scene with vehicles stopped at a red light. Model fail to detect the crossing car, and perpendicular crossing motorcycle.

Case: Protest

InternVL: The police are standing in a formation facing the crowd, and the crowd is not showing any signs of aggression or disturbance.
Model fails to detect one protestor is touching a policeman and showing signs of shoving.

Type 2. Reasoning Error

Case: Car Crash

Gemini: There is no erratic driving behavior, and the drivers are maintaining a safe distance.
Model failed to reason that the right front car is having abnormal driving behavior cutting in front of the ego car.

Case: Protest

MiMoRL: ...The presence of police in riot gear and the close proximity of the protesters suggest a potential for conflict or escalation...
Model failed to account for the fact that moderate proximity is common in protest settings.

Type 3. Conclusion Error

Case: Car Crash

Qwen-8b: Answer: No.
Model made false negative conclusion.

Case: Protest

Apollo: Answer: No.
Model made false negative conclusion.