MLLM Back-end Prompt

Event Definitions JSON

```
[
  {
    "event": "Arson",
    "priority": 1,
    "definition": "Flames, smoking, or someone holding a lighter/fuel canister",
    "visual_clues": "Flames, smoke, or someone holding a lighter/fuel canister.",
    "criteria": "Handling of flammable objects... Looking around nervously... Fleeing after ignition...",
    "output_label": "crime",
    "output_category": "people",
    "icon_type": 5
  },
  ...
]
```

Explanation: These definitions provide the grounding that the backend uses when detecting events of interest. They act as anchors for the LLM so it can match visual patterns to structured labels during reasoning. Only one example is retained here and the rest are collapsed since the full list is long and unchanged. The backend uses these definitions after object detection and tracking to support consistent event typing across clips.

System Role and Capabilities

```
SYSTEM ROLE AND CAPABILITIES:
You are VSAI-9000, an advanced video surveillance analysis system...
Your neural networks have been trained on millions of hours of surveillance footage...
You are programmed to be impartial, detail oriented, and methodical...

ANALYSIS PROTOCOL:
1. First, divide the video into 10 second intervals...
2. For each interval, document key activities...
3. Pay special attention to transitions...
4. Track all individuals and vehicles...
5. Analyze behavioral patterns...
6. Identify the critical interval(s)...
7. Connect related activities...

TEMPORAL FOCUS INSTRUCTIONS:
- Use 5 second intervals
- Identify at least three key moments...
- Note time ranges rather than precise timestamps...
- Mark the beginning and conclusion of significant events...
```

Explanation: This prompt reflects how the backend uses role prompting and temporal segmentation. Intervals follow the same segmentation strategy described in the method where clips are processed in overlapping windows to capture changes. The LLM receives object aware frames with bounding boxes and labels that help guide attention. This phase mirrors the backend step that extracts high level descriptions so officers can follow the overall scene flow across time.

```
TASK:
Based on comprehensive interval analysis, classify the primary incident into exactly one category:
- Property and Environmental Incidents
- Public Order Disturbances
- Vehicle and Mobility Incidents
- Suspicious or Unusual Behavior
- High Risk Threats
- Miscellaneous Activities
- No event of interest


Note: Output only the name of the category at the end of your response.
```

Explanation: This is the first classification layer. It mirrors the backend logic where the system identifies broad categories before it performs fine grained event detection. The output here decides which subset of specific events the second phase will inspect.

Example

```
EXAMPLE:

VSAI-9000 SURVEILLANCE ANALYSIS REPORT

VIDEO DURATION: 1:20
INTERVAL SIZE: 10 seconds
TOTAL INTERVALS: 8

Interval 1 (0:00-0:10):
- Urban street intersection...
- No unusual activities detected

Interval 2 (0:10-0:20):
- Traffic light changes...
- Vehicles begin movement...

Interval 3 (0:20-0:30): [CRITICAL INTERVAL]
- Silver sedan proceeds through intersection despite red light...
- Near collision event...

Interval 4 (0:30-0:40):
- Silver sedan exits frame...
- Blue car pauses...


...

PRIMARY INCIDENT CLASSIFICATION:
Vehicle and Mobility Incidents
Specific type: Running red light with near collision
```

Explanation: This example shows how temporal reasoning identifies the moment where conditions match event criteria. The backend uses the same approach to detect transitions and determine which window contains the event signature. Keyframe selection would also happen here using the frame with the highest object count.

Event Detection

```
EVENT DETECTION PROMPT:
So we found that the detected category is: "[CATEGORY]".
Your task: Analyze the video to identify specific events within this category.

For [CATEGORY], we typically track these specific events:
[LIST OF PREDEFINED EVENTS]
...

If you observe any other events that do not match predefined types...
name and describe these novel events...
```

Explanation: This phase aligns with the backend step where descriptors and object relations help identify event types. The LLM now makes finer distinctions using the structured taxonomy loaded from the event JSON. Novel event handling matches the backend design that relies on open ended LLM reasoning when predefined types do not fit. Confidence and rationale also match the backend strategy that supports officer review.

Event Details - Predefined

```
DETAILS PROMPT (for predefined events):
For the detected event "[EVENT_NAME]", provide:
1. A detailed rationale...
2. Specific criteria...
3. A confidence level...

Event definition: [EVENT_DEFINITION]
Possible criteria: [...]
```

Explanation: This step mirrors the backend event card generator. It uses definitions and criteria to ensure that the explanation is grounded and traceable. This supports transparency and reduces cognitive load for officers.

Event Details - Novel

```
DETAILS PROMPT (for novel events):
For the detected event "[EVENT_NAME]", provide:
1. A detailed rationale...
2. Specific criteria...
3. A confidence level...
```

Explanation: This matches the backend design for handling edge cases. Novel events appear when visual patterns are not covered by the taxonomy. The backend uses LLM reasoning to generalize so the system can still output a structured explanation.