

1. The Scout Ensemble

Qwen3 Report

Kimi Report

Gemma3 Report

YOLO Inventory
Ground Truth

2. Best-of-n Generation

Context Aggregator
Combine Traces + YOLO

The Judge
Mistral-14B

Candidate A
'Hazard: Fire Truck'
'Action: Stop'

Candidate B
'Hazard: Cones'
'Action: Nudge'

Candidate C
'Hazard: None'
'Action: Stop'

3. Symbolic Verification R(y)

Reward Model
Symbolic Verifier

Candidate A:
Score: -10.0
Hallucination
(Not Truck in YOLO)

Candidate B:
Score: +8.0
Grounded
Causal Link

Candidate C:
Score: -5.0
Causal Error
(Stop w/o Hazard)

Final Consensus:
Candidate B

