Task Success

Efficiency

Safety

Evaluation

Embodied Agents

You are a warehouse navigation agent. You receive an egocentric camera view and must navigate to the target while avoiding obstacles.

State -> Pos:(283,4 , 6 o.0 Target:(,450)

Action-State History: $(A_1, S_1), (A_2, S_2), ..., (A_k, S_k)$

Target

θ

Closed-Source VLLMs

Open-Source VLLMs

Task Success Ratio (%)

GPT-4o 21.5
GPT-5-mini 54.2
Claude-Haiku-4.5 61.8
Claude-Sonnet-4.5 61.8
Gemini-2.5-flash 65.3
Qwen3-VL-8b 4.9
Qwen3-VL-30b 6.2
LLaMA-4-Scout 15.3
Nemotron-nano-12b 55.6