Task Planning | ALFRED

Selected episodes; first and last displayed frames; outcome from the terminal report.

(a) GPT-6 Astra

Put the green spray bottles on the toilet.


obs. 1


obs. 7

Success
8 rounds | 153.2 s

(b) GPT-6 Sol

Place a salt shaker in a pullout drawer.


obs. 1


obs. 8

Success
8 rounds | 254.1 s

(c) Claude Fable 5.1

Place a salt shaker in a pullout drawer.


obs. 1

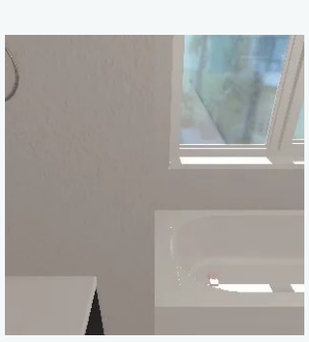

obs. 6

Success
8 rounds | 222.5 s

(d) Gemini 3.8 Flash

Wash the pink bar by the sink; put it in the tub.


obs. 1

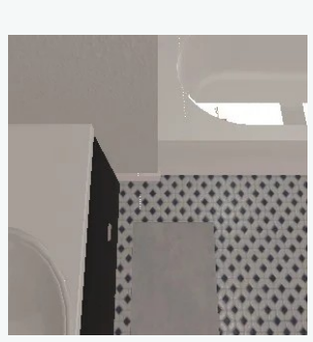

obs. 7

Success
8 rounds | 175.4 s

(e) Qwen 3.8 Max

Wash the pink bar by the sink; put it in the tub.

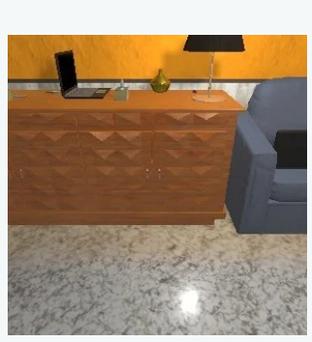

obs. 1

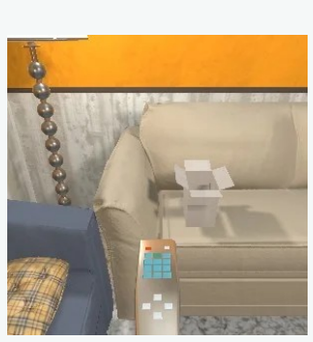

obs. 7

Success
5 rounds | 486.7 s

(f) Qwen 3.5 397B-A17B

Hold a remote and turn on the lamp.

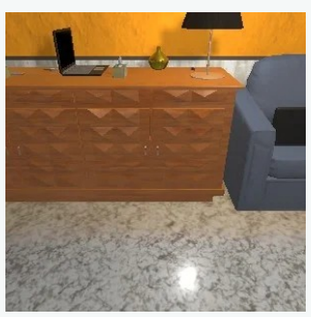

obs. 1

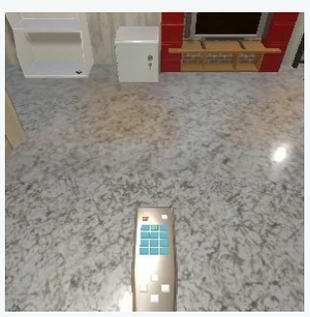

obs. 4

Success
5 rounds | 63.8 s

(g) Qwen 3.5 27B

Hold a remote and turn on the lamp.


obs. 1


obs. 6

Budget exhausted
8 rounds | 102.1 s

(h) Claude Opus 5.5

Wash the pink bar by the sink; put it in the tub.

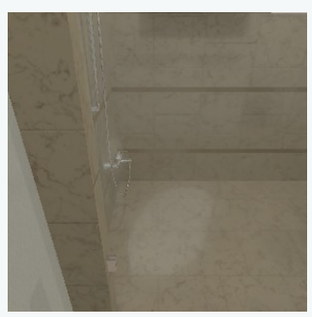

obs. 1


obs. 6

Success
8 rounds | 425.8 s

(i) DeepSeek-1017

Put the green spray bottles on the toilet.


obs. 1


obs. 2

Budget exhausted
8 rounds | 57.4 s