

	Claude Sonnet 4	GLM 4.5V	GPT-5	Gemini 2.5 Pro	Gemma 3 27B Instruct	Grok 4 Fast	Qwen 2.5 VL 72B Instruct	Qwen VL Max	Qwen3 VL 235B Instruct	UI TARS 1.5 7B
Restricted action space and action looping	0.49	0.75	0.70	0.52	0.59	0.67	0.63	0.82	0.62	0.87
State mismanagement	0.29	0.26	0.14	0.33	0.22	0.26	0.17	0.21	0.31	0.35
Failure to use visual or spatial information	0.18	0.25	0.20	0.18	0.13	0.17	0.16	0.27	0.12	0.13
Early termination	0.23	0.19	0.36	0.47	0.34	0.38	0.42	0.17	0.37	0.15