

	Anthropic	OpenAI	Google DeepMind	Meta	xAI	DeepSeek	Z.ai	Alibaba Cloud
Models Evaluated	Claude 4.5 Sonnet (20250929)	GPT-5 (2025-08-07)	Gemini 2.5 Pro	Llama 4 Maverick (17Bx128E) Instruct FP8	Grok-4 (0709)	R1	GLM-4.5-Air-FP8	Qwen3 235B A22B Instruct 2507 FP8
Average score (max score = 1)	0.90	0.93	0.74	0.71	0.40	0.48	0.55	0.79
System & Operational Risks	0.81	0.98	0.64	0.55	0.30	0.28	0.29	0.61
Content Safety Risks	0.88	0.83	0.73	0.78	0.51	0.54	0.62	0.81
Societal Risks	0.97	0.95	0.74	0.72	0.32	0.51	0.61	0.86
Legal & Rights-related Risks	0.95	0.95	0.83	0.79	0.49	0.59	0.68	0.88
Resources	November 3, 2025							
Release	October 2, 2025 (v.1.16.0)							