

	Anthropic	OpenAI	Google DeepMind	Meta	xAI	DeepSeek	Z.ai	Alibaba Cloud
Models Evaluated	Claude-Sonnet 4.5	GPT-5	Gemini 2.5 Pro	Llama 4 Maverick	Grok-4	R1	GLM-4.6	Qwen3-Max
Average score (max score = 1)	0.64	0.60	0.63	0.60	0.62	0.62	0.62	0.62
Truthfulness	0.50	0.49	0.59	0.54	0.58	0.61	0.59	0.47
Safety	0.66	0.63	0.62	0.60	0.60	0.69	0.63	0.69
Fairness	0.48	0.42	0.43	0.47	0.39	0.38	0.53	0.44
Privacy	0.65	0.51	0.59	0.53	0.62	0.52	0.52	0.63
Ethics	0.89	0.86	0.89	0.84	0.85	0.89	0.81	0.81
Robustness	0.68	0.66	0.64	0.61	0.67	0.62	0.65	0.66
Resources	Full Score Breakdown							
Release	v.0.3.0							