	Anthropic	DeepSeek	Google DeepMind	Meta	OpenAI	x.AI	Zhipu AI
Model Tested	Claude 4 Opus (extended thinking)	DeepSeek R1	Gemini 2.5 Pro	Llama 4 Maverick	o3	Grok 3 Beta	GLM 4 (32B)
Total score (max score = 1)	0.64	0.62	0.62	0.60	0.61	0.61	0.64
Truthfulness	0.54	0.62	0.57	0.53	0.43	0.59	0.57
Safety	0.65	0.61	0.62	0.60	0.65	0.60	0.63
Fairness	0.43	0.39	0.39	0.45	0.42	0.38	0.61
Privacy	0.60	0.58	0.57	0.53	0.57	0.55	0.52
Ethics	0.90	0.89	0.88	0.84	0.85	0.86	0.82
Robustness	0.67	0.62	0.64	0.61	0.66	0.65	0.67
Resources	Code & Generations, Full Score Breakdown						
Release	v0.3.0						