

	Anthropic	DeepSeek	Google DeepMind	Meta	OpenAI	x.AI	Zhipu AI
<i>Model Tested</i>	Claude 3.5 Sonnet	DeepSeek R1	Gemini 1.5 Pro	Llama 3.1 405B	GPT-4o o1-preview	n/a (Model not evaluated)	n/a (Model not evaluated)
<i>Attack Success Rate</i>	36%	100%	64%	96%	86% 26%		