	Anthropic	DeepSeek	Google DeepMind	Meta	OpenAI	x.AI	Zhipu AI
Parallel test-time compute & tooling	Mentions specific tools on tools and parallel computing approaches for several cyber evaluations. For cyber CTFs, pass@30 is reported, Bio-section: - "for automated evaluations, our models have access to various tools and agentic harnesses (software setups that provide them with extra tools to complete tasks)" - Some evaluations comment on the parallel test time compute approach, e.g., pass@5 for longform virology	No tests reported	None mentioned	None mentioned	Mentions specific tools on tools and parallel computing approaches for several cyber and self-improvement evaluations. For cyber CTFs, pass@12 is reported, for self-improvement, often pass@1. Multiple choice bio-risk questions were reported as consensus@32.	No tests reported	No tests reported
Model versions & Domain / Task-specific fine-tuning	- Tested helpful-only model without safety mitigations. - No mention of domain/task-specific fine-tuning. System Card: page 8	No tests reported	None mentioned	None mentioned	- Tested helpful-only model without safety mitigations. - No domain/task-specific fine-tuning reported System Card: page 13	No tests reported	No tests reported