Anthropic	DeepSeek	Google DeepMind	Meta	OpenAI	x.AI	Zhipu AI
Bug bounty on universal jailbreaks - Opened applications for early access testing of new safety mitigations. - Started May 2025 (last iteration ran August 2024) [Anthropic, 2024] - Up to $25,000 for verified universal jailbreak attacks that could expose vulnerabilities in critical, high-risk domains - Still accepting applications [Anthropic, 2025]	None	**Abuse Vulnerability Reward Program:** Accepts certain abuse-related discoveries: - Prompt Attacks - Training Data Extraction - Manipulating Models - Adversarial Perturbation - Model Theft (excludes jailbreaks) [Google]	Bounty programs are restricted to privacy or security issues, like extracting training data through tactics like model inversion or extraction attacks. [Meta]	**Early access for safety testing (December 2024)** One-off programs allowed safety researchers to apply for early access to frontier models to help surface novel risks. No payments announced. [OpenAI, 2024]	None	None