

# AI 技术每日分析

中国高技术产业发展促进会新质生产力工作委员会

博雅云创 & 中科创新驱动

2026 年 4 月 21 日

## 摘要

Anthropic Claude Opus 4.7 自 4 月 16 日发布以来持续引发行业深度讨论。The Verge、Axios 等媒体聚焦其“能力选择性释放”策略——Anthropic 首次在模型训练阶段主动实验性削弱网络安全攻击能力，并内置自动化防护栏，为 Mythos 级模型的最终公开发布积累安全经验。中国 AI 公司 MiniMax 于 4 月 12 日开源 M2.7 模型，该模型具备“自我进化”训练能力，在 SWE-Pro 基准上超越 Claude Opus 4.6，推理速度达约 100 tokens/s。Midjourney 发布 V8.1 版本，渲染速度较 V8 提升 3 倍。Stanford 发布 2026 年 AI Index 年度报告，前沿模型在 FrontierMath 上的准确率已从不足 2% 飙升至超过 50%。

## Contents

|     |   |   |
|-----|---|---|
| 1   | <b>Claude Opus 4.7 深度追踪：AI 安全的“能力选择性释放”实验</b> | 2 |
| 1.1 | 首次在训练阶段主动削弱攻击能力 . . . . .                     | 2 |
| 1.2 | 内置自动化网络安全防护栏成行业先例 . . . . .                   | 2 |

|          |   |          |
|----------|---|----------|
| <b>2</b> | <b>MiniMax 开源 M2.7: 自我进化的智能体模型挑战前沿</b>            | <b>3</b> |
| 2.1      | ”模型自我改进”训练范式引发关注 . . . . .                        | 3        |
| 2.2      | SWE-Pro 超越 Claude Opus 4.6, 推理速度达 100 tokens/s    | 3        |
| <b>3</b> | <b>Midjourney V8.1 发布与 Stanford AI Index 2026</b> | <b>3</b> |
| 3.1      | Midjourney V8.1: 渲染速度提升 3 倍 . . . . .             | 3        |
| 3.2      | Stanford AI Index 2026: 前沿模型数学能力一年飙升 25 倍         | 4        |
| <b>4</b> | <b>中国 AI 企业出海与安全博弈</b>                            | <b>4</b> |
| 4.1      | Anthropic 指控中国 AI 实验室大规模”蒸馏”Claude . . .          | 4        |
| <b>5</b> | <b>参考文献</b>                                       | <b>4</b> |

## 1 Claude Opus 4.7 深度追踪: AI 安全的”能力选择性释放”实验

### 1.1 首次在训练阶段主动削弱攻击能力

Claude Opus 4.7 于 4 月 16 日发布后, 本周多家权威科技媒体对其安全策略进行了深度解读。The Verge 引用 Anthropic 原文指出: ”Opus 4.7 是首个此类模型: 其网络安全能力不如 Mythos Preview 先进——事实上, 在训练过程中我们实验性地尝试了差异化削弱这些能力。”这意味着 Anthropic 正在开创一种全新的 AI 安全范式: 不是在模型发布后通过外部过滤器限制危险用途, 而是在训练阶段就主动调控模型的能力边界。Axios 报道称, Anthropic 明确表示将利用 Opus 4.7 在真实世界部署中积累的安全经验, ”帮助我们朝着最终广泛发布 Mythos 级模型的目标迈进”。

## 1.2 内置自动化网络安全防护栏成行业先例

Help Net Security 的分析指出，Opus 4.7 是首个内置自动化网络安全防护栏的商用大模型，能够自动检测并阻止涉及高风险网络安全用途的请求。IT Pro 报道称，Anthropic 在 Opus 4.7 发布中“降低了网络安全能力的调门”，这是对此前 Claude Mythos 在漏洞挖掘方面展现出的惊人能力的直接回应。Benzinga 进一步指出，这种“Mythos 启发的网络安全保护”策略可能成为未来大模型发布的行业标准。值得注意的是，Opus 4.7 在 SWE-bench Verified 上仍达到 87.6% 的成绩，支持 100 万 Token 上下文窗口，商业竞争力并未因安全限制而显著削弱。

## 2 MiniMax 开源 M2.7：自我进化的智能体模型挑战前沿

### 2.1 “模型自我改进”训练范式引发关注

4 月 12 日，中国 AI 公司 MiniMax 正式开源其 M2.7 大语言模型，并在 Hugging Face 上提供下载。M2.7 最引人注目的技术特点是其“模型自我改进”（Model Self-Improvement）训练范式——模型在训练过程中能够自主生成训练数据并进行自我迭代优化。Dataconomy 报道称，M2.7 在软件工程任务上的表现已匹配 GPT-5.3-Codex 的水平，引发了开源社区的广泛讨论。

### 2.2 SWE-Pro 超越 Claude Opus 4.6，推理速度达 100 tokens/s

AI/ML API Blog 的详细评测显示，M2.7 在 SWE-Pro 基准测试上超越了 Claude Opus 4.6，同时推理速度达到约 100 tokens/s，在性能与效率之间取得了出色的平衡。MiniMax 官方将 M2.7 定位为“增强的复杂 Agent 构建能力”模型，强调其在编码和推理方面的行业领先表现。作为中国 AI 企业在开源领域的又一重要贡献，M2.7 的发布进一步丰富了全球开源大模型生态。

### 3 Midjourney V8.1 发布与 Stanford AI Index 2026

#### 3.1 Midjourney V8.1: 渲染速度提升 3 倍

Midjourney 于近日发布 V8.1 版本，在 V8 Alpha (3 月 17 日发布) 的基础上实现了显著的性能飞跃。MindStudio 的对比评测指出，V8.1 相较 V8 实现了 3 倍的速度提升。V8 系列相较 V7 已实现 4-5 倍的渲染加速，标准任务从 30-60 秒压缩至 15 秒以内，V8.1 进一步缩短。此外，V8 系列支持原生 2K 高清图像输出、改进的文字渲染能力和更精准的提示词理解。

#### 3.2 Stanford AI Index 2026: 前沿模型数学能力一年飙升 25 倍

IEEE Spectrum 报道了 Stanford 大学近日发布的 2026 年 AI Index 年度报告。报告中最引人注目的数据之一是前沿模型在 FrontierMath 基准上的表现：2024 年底该基准发布时，最好的模型准确率不足 2%；而截至 2026 年 4 月，Claude Opus 4.6 和 Gemini 3.1 Pro 等前沿模型的准确率已突破 50%。全球 AI 生态系统中已有超过 500 个模型通过商业 API 和开源渠道提供服务。

### 4 中国 AI 企业出海与安全博弈

#### 4.1 Anthropic 指控中国 AI 实验室大规模“蒸馏”Claude

Times of India 报道称，Google、OpenAI 和 Anthropic 三大美国 AI 巨头罕见联手，共同应对硅谷面临的“中国问题”。Anthropic 指控 DeepSeek、Moonshot (月之暗面) 和 MiniMax 等中国 AI 实验室曾使用约 24,000 个欺诈账户对 Claude 进行“工业级规模的知识提取”。三家公司已联合向美国政府发出警告，呼吁加强对 AI 模型知识产权的保护。

## 5 参考文献

1. The Verge (2026 年 4 月 17 日), *Anthropic releases a new Opus model amid Mythos Preview buzz.*
2. Axios (2026 年 4 月 16 日), *Anthropic releases Claude Opus 4.7, concedes it trails unreleased Mythos.*
3. CNBC (2026 年 4 月 16 日), *Anthropic rolls out Claude Opus 4.7, an AI model that is less risky than Mythos.*
4. Help Net Security (2026 年 4 月 16 日), *Anthropic releases Claude Opus 4.7 with automated cybersecurity safeguards.*
5. IT Pro (2026 年 4 月 17 日), *Anthropic toned down Opus 4.7's cyber uses in wake of Claude Mythos release.*
6. Benzinga (2026 年 4 月 17 日), *Anthropic Releases Claude 4.7, Introducing Mythos-Inspired Cybersecurity Protections.*
7. Dataconomy (2026 年 4 月 14 日), *MiniMax M2.7 Matches GPT-5.3-Codex In Software Engineering Tasks.*
8. AI/ML API Blog (2026 年 4 月), *MiniMax M2.7 Review 2026: The Self-Evolving Agentic LLM.*
9. MindStudio (2026 年 4 月 20 日), *Midjourney v8.1 vs MAI Image 2: Which AI Image Model Is Faster in 2026?.*
10. IEEE Spectrum (2026 年 4 月 16 日), *Stanford's AI Index for 2026 Shows the State of AI.*
11. Times of India (2026 年 4 月), *Google, OpenAI and Anthropic come together to fight Silicon Valley's 'Chinese problem'.*

# 联系我们，请扫描二维码



新质生产力工作委员会  
官方公众号



工业智能算网  
gyznsw.cn

## 新质生产力工作委员会：

中国高技术产业发展促进会新质生产力工作委员会，专注于推动工业人工智能、智能制造、数字化转型等前沿技术发展，为企业提供政策解读、技术咨询和产业对接服务。

## 工业智能算网：

专注于工业人工智能、新质生产力、工业软件 CAE、智能制造等前沿技术。提供每日动态分析、技术趋势解读、解决方案分享，推动工业智能化转型。

网站地址：<https://gyznsw.cn>