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 19 日

摘要

国际人工智能领域的焦点高度集中在网络安全风险与大模型在垂直科研领域的能力突破上。最引人瞩目的事件是 Anthropic 尚未公开发布的"Claude Mythos" 模型因其极其强大的漏洞挖掘能力，引发了美国白宫、欧洲央行及全球金融界的强烈震动。同时，Anthropic 发布了 Claude Opus 4.7 版本以及视觉协作工具 Claude Design，进一步巩固了其在复杂推理与代码工程上的优势。作为回应，OpenAI 在垂直科研领域动作频频，不仅推出了专注于生命科学研究的"GPT-Rosalind"，还宣布了新的网络防御生态计划。当前社区讨论热烈，核心争议围绕"AI 漏洞挖掘的防御与攻击失衡" 以及各巨头的安全护城河战略展开。本期分析已排除工业相关内容，主要聚焦基础模型、网络安全及全球政策动态。

Contents

1 核心焦点：Claude Mythos 引发全球网络安全与地缘政治震动

在过去 24 小时及近期的连续发酵中，Anthropic 的未公开模型 Claude Mythos Preview 毫无疑问占据了全球科技媒体和政界讨论的头条。该模型被 Anthropic 官方定义为” 分水岭时刻（Watershed moment）”，因为其在系统级软件漏洞挖掘与利用方面展现出了远超顶级人类黑客的能力。

1.1 突破性的网络安全能力

根据公开的测试报告与业界专家的反馈，Claude Mythos 具备强大的结构化语言（如代码）理解能力，能够发现人类专家及传统自动化工具难以察觉的深层逻辑漏洞。

零日漏洞挖掘：在受控测试中，Mythos 成功发现了开源操作系统 OpenBSD 中潜伏了 27 年之久的底层漏洞，并在视频编码器 FFmpeg 中

找出了此前躲过 500 万次自动化测试的盲点。

时间与效率压缩：传统上，精英安全团队通常需要数周甚至数月才能发掘并编写出针对此类高危漏洞的利用链（Exploit Chain），而 Mythos 将这一过程压缩到了几个小时。

高昂的计算成本：尽管能力惊人，但其运行成本极高。据透露，Mythos 寻找一个存在了数十年的漏洞需要进行数千次运行，单次漏洞发现的算力成本高达约 20,000 美元。

1.2 白宫介入与政界的高压反应

Mythos 的出现直接引发了美国最高决策层的介入。美国东部时间 4 月 17 日（周五），Anthropic CEO Dario Amodei 与白宫办公厅主任 Susie Wiles 进行了紧急会晤。

美政府态度的微妙转变：此前，特朗普政府与 Anthropic 之间曾因五角大楼的 AI 使用协议存在严重分歧。Anthropic 拒绝放宽其使用条款以支持军方某些行动，导致白宫一度试图将其列为”供应链风险”并威胁切断联邦合同。然而，Mythos 在国家安全层面的潜在破坏力，迫使白宫不得不重新回到谈判桌前，以评估该模型对国家基础设施的真实威胁。

Project Glasswing 计划：为缓解风险，Anthropic 宣布暂停对公众开放 Mythos，并联合亚马逊、苹果、谷歌、微软及摩根大通等超 40 家全球顶级企业成立了 Project Glasswing。Anthropic 为该计划提供了 1 亿美元的 API 额度，试图在恶意攻击者获得类似能力前，先利用 Mythos 协助关键基础设施完成全球性的漏洞修补。

1.3 欧洲及全球金融界的恐慌与应对

除了美国，该事件在欧洲及全球金融系统也引发了连锁反应。

英国与欧盟的紧急部署：英国银行业高管在刚刚过去的 24 小时内接到了紧急警告，监管机构正在评估 Mythos 对金融 IT 系统可能造成的”

未知未知（Unknown unknown）”风险。Anthropic 将在下周向部分英国金融机构开放该模型的受控访问权限，以进行防御性测试。

欧洲央行（ECB）表态：欧洲央行行长克里斯蒂娜·拉加德（Christine Lagarde）公开指出，Mythos 的出现展示了负责任的企业如何处理强大的双刃剑技术。她警告称，一旦此类具备自主挖掘系统漏洞能力的 AI 落入黑客或敌对势力手中，对全球经济与公共安全的打击将是毁灭性的。

2 基础模型更新：Claude Opus 4.7 与 Claude Design 齐发

在安全话题占据头条的同时，Anthropic 也在常规的商业与技术落地上发布了重大更新，直接对标 OpenAI 的旗舰产品线。

2.1 Claude Opus 4.7 发布

4 月 16 日，Anthropic 正式推出 Claude Opus 4.7，取代了之前的 4.6 版本，进一步提升了其作为顶级生产力大模型的能力下限与上限。

核心能力提升：Opus 4.7 在高级软件工程、Agent（智能体）执行、视觉处理及多步骤复杂推理任务上表现出了显著的代际提升。许多开发者反馈，以往需要人类密切干预的高难度代码重构任务，现在可以放心地全权交由 4.7 处理。

全新的”xhigh” 努力级别：为了平衡计算延迟与推理深度，Anthropic 引入了介于 high 和 max 之间的新参数设置 xhigh（Extra High）。这一设置将赋予用户更精细的控制权，允许模型在面对复杂逻辑时投入更多算力思考。在 Claude Code 产品中，默认的计算级别已被集体上调至 xhigh。

底层优化与计费调整：Opus 4.7 采用了全新的分词器（Tokenizer），对文本的处理更加高效，但这导致相同文本的 Token 消耗量增加了约 1.0 至 1.35 倍。尽管 API 定价维持不变（输入 $5/1M Tokens，输出 $25/1M

Tokens)，但实际使用成本有小幅上升。同时，新版本内置了更为严格的网络安全防护栏，以自动阻断高危代码的生成请求。

2.2 视觉协作新产品：Claude Design

4 月 17 日，Anthropic Labs 推出了 Claude Design。这是一款全新的生产力工具，旨在让用户与 Claude 进行沉浸式的视觉协作。用户可以直接通过对话生成、修改和打磨高保真原型图、幻灯片设计以及单页宣发物料（One-pagers)。此举标志着 Anthropic 正加速侵入传统的 UI/UX 设计与办公软件赛道。

3 OpenAI 的垂直突围：GPT-Rosalind 与网络防御生态

面对 Anthropic 在安全与底层推理上的咄咄逼人，OpenAI 在过去 24 小时内选择了" 垂直深度" 与" 生态防御" 作为破局点。

3.1 GPT-Rosalind 进军生命科学

OpenAI 于 4 月 16 日正式发布了 GPT-Rosalind（名称致敬 DNA 双螺旋结构发现先驱 Rosalind Franklin)。这是一个专门针对生命科学研究微调与优化的大型语言模型。与通用大模型不同，GPT-Rosalind 在基因组学、蛋白质折叠原理分析、制药文献合成等生物学前沿领域具有极高的准确率和幻觉抑制能力。这标志着 OpenAI 正在将战火从通用 AGI 烧向需要极高专业壁垒的垂直科研领域。

3.2 强化网络防御生态

为了呼应当前业界对 AI 加剧网络攻击的担忧，OpenAI 同步发布了主题为 *"Accelerating the cyber defense ecosystem that protects us all"* 的战略更新，并配合推出了新一期的 OpenAI 安全研究人才奖助计划。OpenAI 强调，抵御 AI 攻击的最好方式是建立基于 AI 的开源与闭源协同防御网络，试图在话语权上与 Anthropic 的" 闭门造车（Project Glasswing）" 形

成对比。

4 社交媒体与技术社区热点

在 Reddit（如 r/LocalLLaMA, r/MachineLearning）和 Twitter/X 的技术圈内，过去 24 小时的讨论几乎全部被”Mythos 的安全争议” 与”Opus 4.7 的实测” 占据。

” 伪善” 还是” 真负责”？许多开源社区的支持者对 Anthropic 的做法表示了强烈的冷嘲热讽。Reddit 上的高赞评论指出：”Anthropic 上个月才刚意外泄露了部分 Claude 内部源代码，这个月就突然声称自己是全球网络安全的’ 守护者’，甚至要把新模型锁起来不让大家用。如果换作是其他大厂，早被嘲笑透顶了。”

矛与盾的成本博弈：安全研究人员在 Twitter 上热烈讨论了 Mythos 单次两万美元的漏洞挖掘成本。共识认为，虽然短期内只有大型科技公司或国家级黑客（Nation-states）能承担这种级别的攻击/防御成本，但随着算力的下降，这种能力必将下放。这引发了人们对”AI 时代防御将永远落后于攻击” 的深层焦虑。

Opus 4.7 的落地反响：在实测反馈中，开发者普遍对 Opus 4.7 的重构能力表示惊艳，尤其是配合全新的 xhigh 思考模式，模型在处理包含数千行陈旧代码的代码库时，展现出了极强的上下文一致性。

5 参考文献

1. The Washington Post (2026 年 4 月 17 日), *Anthropic CEO visits White House amid hacking fears over new AI model.*
2. The Guardian (2026 年 4 月 17 日), *Finance leaders warn over Mythos as UK banks prepare to use powerful Anthropic AI tool.*
3. PBS NewsHour (2026 年 4 月 17 日), *White House chief of staff to meet*

with Anthropic CEO over its new Mythos AI model.

4. Platformer (2026 年 4 月 8/17 日), *Why Anthropic's new model has cybersecurity experts rattled.*
5. AP News (2026 年 4 月 18 日), *White House chief of staff meets with Anthropic CEO over its new AI technology.*
6. Anthropic Newsroom (2026 年 4 月 16 日), *Introducing Claude Opus 4.7.*
7. Anthropic Newsroom (2026 年 4 月 17 日), *Introducing Claude Design by Anthropic Labs.*
8. The Guardian (2026 年 4 月 12 日), *'Too powerful for the public': inside Anthropic's bid to win the AI publicity war.*
9. Times of India (2026 年 4 月 10 日), *Explained: Why Anthropic's Claude Mythos is scaring the company so much that it has decided to not release it to public.*
10. OpenAI Newsroom (2026 年 4 月 16 日), *GPT-Rosalind for Life Sciences & Accelerating the cyber defense ecosystem.*

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