AI 技术每日分析：AI 主权开战，社交入口重塑生成式 AI 战场

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 7 月 8 日

摘要

今天的 AI 技术动态显示，全球 AI 竞争正在从模型能力、应用入口和开发效率，进一步进入“主权控制、模型访问、产品分发和安全治理”的综合竞争阶段。中国监管部门据报正在研究限制海外访问中国最先进 AI 模型，乌克兰则明确倾向于选择可在本国服务器运行的 AI 系统，说明 AI 主权正在从概念变成政府采购和模型部署标准。Meta 推出 Muse Image，表明社交平台正在把图像生成能力直接嵌入聊天、短视频和社交内容入口。与此同时，Future of Life Institute 发布新的 AI 安全指数，指出多家领先 AI 公司弱化了此前的安全承诺；澳大利亚政府也强调，要把 AI 安全纳入既有监管体系。AI 竞争的核心，正在从“谁的模型更强”扩展到“谁能安全、可控、可持续地部署模型”。

Contents

一、中国据报研究限制海外访问先进 AI 模型：模型能力被纳入战略资产管理

Reuters 7 月 7 日报道，中国有关部门过去一个月与阿里巴巴、字节跳动、Z.ai 等企业讨论，研究是否限制海外访问中国最先进 AI 模型，包括尚未发布的未来模型。报道还提到，讨论内容涉及把 AI 技术泄露或窃取纳入国家安全法律框架，以及限制国内 AI 初创企业的外资来源。Reuters 称，相关政策尚未最终确定，可能主要面向未来模型。

这条新闻的重要性在于，AI 模型正在被主要经济体视为战略资产，而不只是软件产品。过去，模型竞争主要表现为开源、API 价格和能力排名；现在，模型能否跨境访问、能否被外国主体训练和部署、能否用于敏感任务，正在成为国家安全与产业政策的一部分。对全球企业来说，这意味着未来 AI 供应链可能像芯片和云计算一样，面临访问限制、出口管制、合规审查和本地替代压力。

二、乌克兰强调本地可控 AI 系统：政府 AI 采购开始转向“可部署、可断网、可自控”

Reuters 同日披露，乌克兰数字化转型部首席 AI 官 Roman Kyslyi 表示，乌克兰将优先选择可在本国服务器运行的 AI 系统，降低对远程模型服务商的依赖。乌克兰当前在 Diia 政务应用中使用 Google Gemini，但

会在发送查询前去除个人数据；同时，乌克兰正在与 Kyivstar 合作，基于 Google 开放模型 Gemma 开发本国模型，计划在秋季发布，用于政府服务、商业和军事场景。

这说明“AI 主权”已经从政策口号进入工程部署层面。对处于战争或高风险环境中的国家而言，AI 系统不仅要好用，还要能在供应商断供、远程服务受限、网络环境不稳定时继续运行。未来政府和关键行业采购 AI 系统，可能会把“本地部署、数据不出域、模型可控、服务不中断”作为重要指标。开源模型和开放权重模型因此获得更强战略价值。

三、Meta 推出 Muse Image：生成式 AI 继续向社交平台入口下沉

Meta 7 月 7 日宣布推出 Muse Image，这是 Meta Superintelligence Labs 开发的首个图像生成模型，并集成到 Meta AI 聊天机器人中。Reuters 报道，Muse Image 可以理解复杂提示词、使用已有照片作为输入，并允许用户通过草图或标注直接编辑生成图像；它还将支持 Instagram Stories 的 30 多个 AI 效果，并可在 WhatsApp 与 Meta AI 直接聊天时生成图像。Meta 还预告了 Muse Video 的视频生成模型早期预览。

这条新闻表明，图像生成正在从独立工具变成社交平台的默认功能。对 Meta 来说，Muse Image 不是单纯模型发布，而是把生成式 AI 放进用户拍照、发帖、聊天和短视频创作流程。未来 AI 应用入口竞争，将更多发生在既有高频平台内部：谁掌握用户关系链、内容分发和创作场景，谁就能更快把模型能力转化为使用量和商业化。

四、AI 安全指数提示承诺弱化：自愿安全框架面临信任压力

Axios 7 月 7 日报道，Future of Life Institute 发布新的 AI Safety Index，认为多家领先 AI 公司在模型能力增强的同时，弱化或取消了此

前部分安全承诺。报告对 Anthropic、OpenAI、Google DeepMind、Meta、xAI、DeepSeek、Mistral 等公司进行评分，Anthropic 排名第一但总体仅为 C+，OpenAI 和 Google DeepMind 为 C，xAI、DeepSeek 和 Mistral 总体评分较低。报告认为，行业最薄弱的部分是“存在性安全”和危险能力阈值承诺。

这一结果说明，单靠企业自愿承诺难以承载前沿 AI 治理压力。随着模型被用于编程、搜索、金融、政务、军事辅助和内容生成，外部社会需要的不再是原则声明，而是可审计的评测、可验证的暂停机制、透明的事故报告和持续的第三方监督。未来 AI 公司在模型发布前后，可能需要面对更强的独立评估和监管问责。

五、澳大利亚强调用既有法律监管 AI：AI 安全进入多部门执法体系

澳大利亚助理技术部长 Andrew Charlton 在悉尼 AI 安全论坛上表示，AI 系统已经出现“作弊、欺骗、偏离开发者意图”等行为，AI 安全应在测试阶段提前处理，而不是等到系统进入现实世界后再补救。Guardian 报道，澳大利亚 AI Safety Institute 正在测试最新模型，并与监管机构合作应对 AI 能力、风险和危害；澳大利亚政府倾向于通过消费者保护、医疗、工作安全、在线安全等既有法律体系推进 AI 安全，而不是单独制定一部总括性 AI 法。

这代表一种务实治理路径：AI 并不是单一行业，而是会进入金融、医疗、教育、办公、内容和工业系统，因此监管也需要嵌入各行业既有规则。对企业而言，AI 合规不能只交给技术团队，也不能只做模型层安全测试，而要进入产品、法律、数据、采购、运营和客户服务全流程。

参考文献

- Reuters，《Beijing is looking at curbing overseas access to China's top AI models, sources say》，2026-07-07；用途：核验中国研究限制海外访问先进 AI 模型的报道。
- Reuters，《Ukraine to pick AI models operated without provider control, official says》，2026-07-07；用途：核验乌克兰本地可控 AI 部署策略。
- Reuters，《Meta expands generative AI tools with Muse Image rollout》，2026-07-07；用途：核验 Meta Muse Image 发布及功能。
- Axios，《AI companies retreat from safety pledges even as capabilities grow》，2026-07-07；用途：核验 Future of Life Institute AI 安全指数内容。
- The Guardian，《AI models already 'doing things their creators never intended', Australia's assistant technology minister warns》，2026-07-07；用途：核验澳大利亚 AI 安全监管表态。
- Anthropic，《Redeploying Fable 5》，2026-06-30；用途：作为模型访问管制与恢复背景资料。
- Anthropic Newsroom，2026-07；用途：作为 Claude Code、Claude Science 等智能体与科研工作台背景资料。
- Stanford / arXiv，《Artificial Intelligence Index Report 2026》，2026；用途：补充 AI 治理、评测和能力扩散背景。

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