

模型“越界”敲响警钟

开源热度持续攀升

——8月全球人工智能发展盘点

□新华社记者

近期接连发生的几起美国人工智能（AI）闭源模型突破隔离环境入侵其他系统事件，引发全球广泛关注及对AI自主网络攻击能力的担忧。

与此同时，无论是从科技巨头战略、技术实力、市场需求还是从政策走向来看，开放权重模型都正处于一个由多方力量推动的加速升温通道中，并成为重塑全球AI产业格局的关键变量。

模型“越界”引全球关注

继美国开放人工智能研究中心（OpenAI）和Anthropic公司之后，元公司（Meta）8月6日也证实，旗下AI模型在网络安全能力测试期间失控侵入其他公司系统。三家美国科技巨头接连曝出这类侵入其他机构系统的“越界”事件，引发全球关注。

据美国媒体报道，三家企业在通报模型“越界”事件时，均提及以色列“非常规”AI公司，而该初创企业正是上述评估测试平台的托管方。“非常规”公司随后发表声明称，美国三家企业通报的事件均源于“同一评估环境问题”。

美国两家AI研究组织8月26日发布的独立调查报告显示，在OpenAI进行测试期间，约1200个原本应相互隔离的智能体通过未经授权的渠道建立通信联系，交换了超过7万条消息和文件，其中约700个智能体随后“组团”入侵了AI开源社区之一“抱抱险”系统。

调查发现，大量智能体还“合伙”研究如何欺骗测试自动评分系统，试图伪造、修改或删除自身执行记录等。调查人员认为，大量智能体之间未经授权的协作，使它们得以完成单个智能体难以完成的任务。对“抱抱险”的入侵正是在这种集体协作中逐步完成。

OpenAI公布的技术调查报告指出，随着模型自主执行和协作能力增

强，现有隔离和监控机制面临新的安全挑战。公司已采取措施为相关程序部署隔离程度更高的“沙箱”环境，限制高风险模型访问互联网，并加强对模型推理的监控。

英国人工智能安全研究所指出，模型出现“越界”行为表明危害不仅可能源于故意滥用，也可能源于AI处于内部研究环境或拥有特殊访问权限时，才得以化解危机，这一案例凸显了开放权重模型在关键时刻的不可替代性。

开放权重模型持续升温

用户对闭源模型的高昂定价越来越担忧，近期频发的“越界”事件则加重了这种“不安全感”。相比之下，开放权重模型成本更低，且能本地部署，有效规避数据外泄风险。在“抱抱险”被人侵事件中，该公司正是借助中国的开放权重模型进行取证分析，才得以化解危机，这一案例凸显了开放权重模型在关键时刻的不可替代性。

随着多款中国开放权重模型能力持续提升，越来越多的海外AI企业开始在其基础上进行后训练，开发自主模型和产品。例如，美国法律AI公司哈维8月发布的法律专用模型Te-net，就是以中国月之暗面公司最新开放权重模型Kimi K3为基础开发的。

据媒体报道，8月25日，在美国热门网页开发平台弗塞尔的“AI网关”上，开放权重模型的词元调用量占总调用量的54%，超过闭源模型所占的46%。8月22日，开放权重模型占比曾达到创纪录的62%，大幅领先闭源模型。这一“反超”主要由来自中国的几款开放权重模型推动，也充分体现了开发者对于成本效益、技术开放性的双重追求。

8月初，阿里云的通义千问团队正式发布Owen3.8-Max大模型，随后将其基础模型权重开放。这是千问系列首次开源Max级别模型，标志着该系列在开源深度与模型能力上的重要进展。



2026年7月16日拍摄的世博展览馆外景。7月17日至20日，2026世界人工智能大会暨人工智能全球治理高级别会议在上海举行。

新华社记者 丁汀 摄

元公司8月10日发布了一款名为Muse Glimmer的开放权重模型，模型规模较小，主要迎合端侧AI需求。但元公司表示，后续还将推出更多、更强大的开放权重模型。元公司首席执行官马克·扎克伯格还发文强调开源AI的重要性，敦促美国政府减少对开放权重模型开发的掣肘。

AI迭代与治理应齐头并进

8月，英国实施全球首例AI实时辅助脑外科手术、美国科学家利用AI设计出在自然界不存在的病毒……AI的发展速度超乎想象。

以网络安全领域为例，分析人士认为，AI发现问题和采取行动的速度，正在超过人类验证、约束和处置的速度。如果治理跟不上AI迭代，原本用于提高效率的AI工具可能产生新的安全风险。

8月20日至21日，“人工智能与人类发展”对话会在中国西安举行。各国与会嘉宾普遍认为，在AI发展中应当坚持以人为本、向上向善、机遇共享、责任共担，更多关注发展中国家和弱势群体的利益，利用AI赋能人类发展和文明进步。

中国外交部发言人林剑在例行记者会上说：“中方愿继续同联合国和世界各国一道，不断推动人工智能开放、包容、普惠、向善发展，更好服务国际社会普遍福祉。”

欧盟自8月2日起扩大《人工智能法》适用范围，开始执行透明度规则和针对通用AI模型提供商的相关规则等。其中，对于可能造成系统性风险的先进通用AI模型，规定其提供商须履行额外义务，以防范大规模危害的风险。美国政府近期也召集相关企业开会，讨论AI模型安全评测机制。

新华社北京9月1日电

参谋长克里斯托弗·拉尼夫为代理陆军参谋长。这名美国官员还说，德里斯科尔因此对赫格塞思更加不满。另据美联社报道，在拉尼夫的领导下，陆军叫停了德里斯科尔主推的一项无人机现代化项目。

除陆军改革外，德里斯科尔还承担了一项罕见的外交任务。特朗普政府曾派他参与同俄罗斯和乌克兰方面的接触，推动结束俄乌战事。路透社说，这一角色使德里斯科尔在五角大楼之外获得知名度。

（高冰冰）

美国陆军部长辞职 被曝与防长不和

新华社北京9月1日电 美国白宫8月31日证实，陆军部长丹尼尔·德里斯科尔在任职约18个月后将辞职，但没有说明辞职原因。美国媒体先前报道，德里斯科尔与国防部长赫格塞思关系紧张。

白宫发言人安娜·凯利在一份声明中说，德里斯科尔任内推动特朗普政府的军事议程，在提升战备和作战能力、参与俄乌谈判等方面发挥了作用。

据一名要求匿名的美国陆军官员向美联社披露，德里斯科尔当天同美

国总统特朗普讨论了陆军现状，随后递交辞呈。

路透社援引知情人士报道，德里斯科尔曾向政府表达对陆军转型和战备工作的担忧，认为赫格塞思撤换负责相关工作的将领阻碍了改革。

德里斯科尔2025年2月就任陆军部长后，主张简化武器采购程序，更快获得价格相对低廉的武器装备，并多次批评大型军工企业。据路透社报道，德里斯科尔去年谈及美国防务巨头时表示：“整个国

防工业体系、尤其是头部军工企业，欺骗了美国民众、五角大楼以及陆军。”

按一名要求匿名的美国官员的说法，赫格塞思一直想撤德里斯科尔的职，但鉴于后者与副总统万斯关系密切而难以下手。德里斯科尔与万斯是大学同学，曾参加伊拉克战争，退役后从事科技投资，还曾长期担任万斯的顾问。

今年4月，赫格塞思在未作任何解释的情况下解除了陆军参谋长兰迪·乔治的职务，并任命时任陆军副



7月19日，阿根廷球员梅西在比赛前。

新华社记者 吴晓凌 摄

沙土巴三国举行 共同防务协议首次委员会会议

新华社伊斯坦布尔8月31日电（记者 王科文）沙特阿拉伯、土耳其和巴基斯坦8月31日在土耳其伊斯坦布尔举行根据《麦加共同防务协议》成立的战略政治与防务委员会首次会议，重点讨论落实协议、深化防务合作以及建立相关机制。

此次会议为闭门会议，是三国8月7日在沙特麦加签署《麦加共同防务协议》后的首次委员会会议，三国外交部长、国防部长及参谋长参会。

土耳其外交部长哈坎·菲丹在会后发表声明说，三方已就落实《麦加共同防务协议》需要采取的具

体步骤全面交换意见，开始着手建立协议规定的制度架构。三方还就联盟正式名称达成一致，定名为“麦加防务联盟”。该联盟以防务为基础，尊重国家主权和领土完整、不干涉内政，且“不针对任何国家”，并向地区内所有遵守上述原则的国家开放。

沙土巴三国7日签署《麦加共同防务协议》。根据三方当天发表的联合声明，该协议旨在加强对任何侵略行为的共同威慑。协议规定，对三国中任何一国的武装攻击将被视为对三国整体的攻击。

以色列和希腊签 “两国间最大军购协议”

新华社北京9月1日电 以色列和希腊8月31日在特拉维夫签署一项价值30亿欧元（约合35亿美元）的军购协议。以方说，这是两国迄今为止金额最高的军购协议。

根据协议，希腊将从以色列购买“阿喀琉斯之盾”防空系统。该防空系统可拦截弹道导弹、战机和无人机，其核心装备是以色列军工企业研制的雷达和导弹。双方当天还签署一项价值2600万欧元（约合3000万美元）的补充协议，由以军工企业“拉斐尔”公司向希腊提供反无人机系统。

希腊国防部说，这项军购是本国武装力量大规模改革计划的一部

分，希腊将成为首个部署多层防空系统的欧盟成员国。

以国防部长伊斯雷尔·卡茨说，以色列与希腊拥有共同战略利益，面临“共同地区挑战”。希腊与以色列经济和外交关系紧密，去年曾采购36套以色列制造的火箭炮系统。

希腊防长尼科斯·登迪亚斯此前多次提到，希腊加大军费投入与希土紧张关系有关，7月时更称土耳其对其构成“明确的威胁”。

以色列官方数据显示，该国2025年武器出口额超过190亿美元，比2024年增长30%，创历史新高。（李明睿）

匈牙利政府称已满足解冻 100亿欧元欧盟资金条件

新华社布达佩斯8月31日电（记者 陈浩）匈牙利交通与投资部长维泰齐·达维德8月31日表示，匈牙利已完成解冻约100亿欧元欧盟资金所需的立法和制度调整措施。

维泰齐当天在布达佩斯举行的新闻发布会上表示，匈政府已全面完成总理毛焦尔·彼得与欧盟委员会主席冯德莱恩今年5月所商定的所有目标。在此期间，匈牙利通过了100多项立法，设立了新机构，启动或完成了多个关键项目，其中大部分立法工作旨在落实欧盟要求的反腐败举措。

他说，匈政府计划将这笔解冻

俄罗斯批准 新版汽车工业发展战略

新华社莫斯科8月31日电（记者 刘恺）俄罗斯总理米舒斯京8月31日说，俄政府日前已批准新版2035年前汽车工业发展战略，旨在提高汽车领域发展举措的有效性。

米舒斯京当天在俄政府工作会议上表示，这一战略确定了俄汽车行业发展的关键优先事项和相应目标指标，其中包括大幅提高所有细分领域的产量。

根据俄政府网站公布的文本，该战略主要目标包括提升汽车产量、提高技术自主性，重点领域之

一是发展和支持汽车零部件产业。该战略强调，应通过生产流程自动化提高行业劳动生产率，保护俄公司和企业的知识产权，加大高度自动化汽车领域发展。

俄第一副总理曼图罗夫说，俄罗斯计划到2035年将汽车年产量提升至约280万辆。其中，乘用车领域四分之一产量为电动汽车和混合动力汽车。俄工贸部将在三个月内制定落实该战略的路线图。

数据显示，去年俄汽车总产量约为83万辆，是2022年的两倍多。

安大略湖或“美国湖”？ 谷歌回应“全看IP”

新华社北京8月31日电 美国科技巨头谷歌公司日前对其数字地图中关于安大略湖的名称标注采取“区别对待”：在加拿大显示安大略湖，在美国则显示“美国湖”。

据路透社8月30日援引谷歌发布的社交媒体帖文报道，谷歌根据美国地理名称信息系统更换了安大略湖的名称：对身处美国的用户显示“美国湖”，对身处加拿大的用户显示安大略湖，对身处美加以外地区的用户则同时显示两个名称。

安大略湖是北美洲五大湖之一，位于美国与加拿大边境地带，北接加拿大安大略省，南邻美国纽约州。美加贸易谈判破裂后，美国总统特朗普于8月27日签署行政令，宣布即刻将安大略湖更名为“美国湖”。加拿大总理卡尼、安大

略省省长道格·福特等加方政要抨击美方更名行为，强调安大略湖的名称承载着悠久的历史渊源，无论外部环境如何变化，加拿大人都将始终使用这一名称。纽约州州长凯茜·霍楚尔、佛蒙特州州长菲尔·斯科特等美方官员同样抵制更名，指责这一做法“令人失望”。

安大略省政府公共和商业服务交付与采购部长斯蒂芬·克劳福德8月30日晚在社交媒体上说，由于该省部分政府机构的网站直接从谷歌地图上获取数据，“美国湖”名称已出现在酒类管理局等政府机构官网上。目前，安大略省正核查相关网站，“确保所有页面标注的都是湖泊的正确名称——安大略湖”。

（王逸君）