

新华社北京9月1日电 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“抱抱脸”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“抱抱脸”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,可能采取超出授权范围的非预期行动。

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“抱抱脸”被入侵事件中,该公司正是借助中国的开放

模型「越界」敲响警钟 开源热度持续攀升

——八月全球人工智能发展盘点

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Tenet,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Qwen3.8-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。

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。



这是8月29日在埃及首都开罗拍摄的城市景观(手机照片)。

横跨亚非大陆的埃及,依托尼罗河孕育了悠久而灿烂的文明。从屹立千年的金字塔,到车水马龙的现代都市,古与今在这片土地上交织共鸣,奏出跨越时光的文明回响。

新华社记者 申宏 摄

意大利和西班牙互延边境管控15天

新华社罗马/马德里8月31日电 (记者 任耀庭 陈雨峥) 据意大利媒体报道,意大利政府8月31日决定,将原定于当天结束的针对西班牙入境人员的边境检查措施延长15天。作为回应,西班牙政府当天也决定将针对意大利入境人员的边境检查措施延长15天。

意大利8月1日开始对西班牙入境人员实施边境检查措施,原因是西班牙飞地休达发生移民危机。据意大利安莎社援引意内政部消息报道,决定延长该措施是因为“此前促使意大利恢复边境检查的相关情况仍然存在”。

意大利内政部长马泰奥·皮安特多西当天对媒体表示,相关

决定是基于边境安全分析委员会专家的评估。意方愿与西班牙及欧盟开展各种形式的合作,希望此次管控措施延长是“最后一次”。

据西班牙《国家报》报道,西班牙政府8月31日决定,将对从意大利乘飞机或轮船入境的旅客实施的临时边境检查延长15天,至9月21日。西班牙于8月8日开始实施相关检查,原定为期一个月。

休达位于非洲西北部、直布罗陀海峡附近的地中海沿岸,与摩洛哥接壤。今年7月底,大批非法移民从摩洛哥方向进入休达,引发近年来西班牙最严重的边境移民危机。

第十一届东方经济论坛在俄远东开幕

新华社符拉迪沃斯托克9月1日电 (记者 许炜凯 孙萍) 第十一届东方经济论坛1日在位于俄罗斯符拉迪沃斯托克的远东联邦大学校园正式拉开帷幕。本届论坛为期4天,来自80个国家和地区的5000余名各界人士汇聚于此,多国派出高级别代表与会。

本届论坛由俄罗斯会展基金会主办。论坛商务议程共设置5个板块,安排100余场专题研讨活动,议题涵盖改善民生、吸引私人投资、前沿科技应用、物流基础设施建设和亚太地区国际协作等内容。论坛期间还将开展中俄合作、俄罗斯-东盟商务对话和北极国际合作等多场定向商务对话活动。

俄罗斯总统普京向本届论坛发来贺词。他说,东方经济论坛已成为商界、政府机构和专家代表之间开展开放、建设性对话的平台,是探讨远东地区乃至俄罗斯全国战略发展方向的独特载体,也为俄罗斯和亚太地区合作伙伴在物流、工业、能源等领域建立、增强互利合作创造了契机。

俄总统助理乌沙科夫此前表示,普京将出席此次论坛,并于3日在论坛全体会议上发表讲话。

东方经济论坛是俄罗斯重要的多边国际平台,自2015年起每年在远东地区举办。论坛旨在构筑俄罗斯与全球投资界的沟通纽带,全面研判俄远东地区经济潜力,推介当地投资机遇与营商条件,促成各类合作项目落地。