

AI 模型越界敲响警钟



智能治理全球对话「现场」(资料片)
7月6日,在瑞士日内瓦,参会者在首届「人工

新华社

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,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“抱抱险”被入侵事件中,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。

新华社北京9月1日电



智元远征A3机器人在展会现场打乒乓球。(资料片) 新华社

美国发生“随机无端袭击”

纽约时报广场持刀行凶事件致1死1伤

新华社纽约8月31日电 据美国纽约市警方消息,一名女子8月31日下午在纽约时报广场持刀行凶,造成一人死亡,一人受伤。嫌疑人已被警方击毙。纽约市长马姆达尼和警方当

晚在新闻发布会上确认,一名受害者伤重不治,另一人伤势稳定。纽约市警察局局长杰茜卡·蒂施将这起事件定性为“随机且无端的袭击”。警方已封锁周边区域道路,调查仍在进行中。

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。

美国向伊拉克出售武器

称助力其应对当前及未来各类安全威胁

新华社北京9月1日电 美国国务院8月31日宣布,已批准向伊拉克出售总额约8亿美元的多种用途直升机。

美国国务院在一份声明里说,这项军售的主要承包商为总部位于美国得克萨斯州沃思堡的贝尔直升机德事隆公司,计划向伊拉克方面出售贝尔412EPX直

升机、贝尔407M直升机等装备,并提供相应的后勤及配套支持。

声明说,伊拉克政府要求购买直升机、机枪架、火箭发射器等其他多种设备。此次军售将提升伊军方旋翼机运输、侦察以及空地目标打击能力,助力伊拉克应对当前及未来各类安全威胁。

美债收益率走高

升至去年1月以来高点

新华社纽约8月31日电 受国际油价走高、市场预期美联储将加息抑制通胀等因素推动,美国10年期国债收益率8月31日明显走高,升至2025年1月以来高点。

数据显示,美国10年期国债收益率当日上涨3.9个基点,收于4.756%,为2025年1月15日以来最高水平;美国30年期国债收益率上涨3.7个基点,收于5.244%。

美国30年期国债收益率8月17日达5.31%,创2007年6月以来新高。外界认为,这促使美国财政部近日扩大了长期国债回购规模。

美国联邦政府债务总额8月首次突破40万亿美元,本财年利息支出预计接近1.2万亿美元。债务规模不断扩大,而高利率又推高偿债成本,美国财政正面临越来越沉重的融资压力。

欧元区8月通胀率明显上升

升至3.3%

新华社布鲁塞尔9月1日电 欧盟统计局9月1日公布的初步数据显示,2026年8月欧元区通胀率按年率计算为3.3%,较7月的2.9%明显上升。

数据显示,8月欧元区能源价格同比上涨14.3%,涨幅较7月的10.3%明显扩大,是推高整体通胀率的主要因素;服务价格同比上涨3.0%,低于7月的3.3%;非能源工业产品价格同比上涨1.2%,高于7月的0.9%。食品、酒类 and 烟草价格同比上涨1.2%,涨幅与7月持平。

剔除能源、食品、酒类和烟草价格后,8月欧元区核心通胀率为2.4%,较7月的2.5%略有回落。

从国别看,8月欧元区主要经济体通胀率普遍回升。其中,德国通胀率从7月的2.8%升至2.9%,法国从2.4%升至2.7%,意大利从2.9%升至3.2%,西班牙从3.9%升至4.5%。

欧洲中央银行理事会成员、芬兰央行行长奥利·雷恩日前警告,面对当前通胀压力,欧洲央行“绝不能掉以轻心”。

日本国债收益率创新高

近30年来首次升至3%

新华社东京9月1日电 日本10年期国债收益率1日一度升至3%,创近30年来新高。

日本相互证券公司当天指出,这是日本10年期国债收益率自1996年10月以来首次升至3%。

美国财政部长贝森特8月31日说,他预计日本政府和日本央行将采取措施推动日元走强。这

一表态增强了市场对日本央行9月加息的预期,成为日本国债收益率上行的重要推手。

中东局势恶化带来的原油价格上涨,同样形成利率上行压力。日本能源高度依赖进口,市场预判日本国内通胀会加剧,进而抛售国债。此外,市场还担忧日本政府财政状况恶化,也加剧了日本国债抛售压力。

指路

9月1日,在新加坡小印度地铁站内,一款名为“Olly”的人形机器人向乘客指路。

8月31日至9月11日,新加坡陆路交通管理局在地铁站测试使用人形机器人“Olly”。乘客可向其询问如何在车站内通行以及地铁线路等问题。

新华社

儿童对极端高温尤其敏感,其体温上升速度快于成年人,出汗散热能力也相对较弱

气候变化损害儿童健康

新华社赫尔辛基9月1日电 东芬兰大学和比利时布鲁塞尔自由大学等机构完成的一项最新研究显示,全球40%以上的10岁以下儿童每年因人为造成的气候变化而多经历至少30天的危险高温。

相关研究成果已发表在美国《科学进展》杂志上。东芬兰大学发布新闻公报介绍,研究团队聚焦湿热天气对各年龄段人群的影响,将人口数据与先进气候模型以及气候归因科学相结合。湿热比干热更危险,因为湿热会阻碍身体通过出汗降温,从而导致

危险水平的热应激。

热应激是指在湿热环境下,当湿球黑球温度超过28摄氏度时,人体因无法有效散热而产生的生理压力状态。研究显示,在当前全球变暖水平下,全球约5.6亿0至9岁儿童每年因人为造成的气候变化而多经历至少30天热应激,约占这一年龄段儿童总数的43%。这一热暴露水平高于其他年龄段人群,凸显年轻一代正面临更重的气候变化负担。

研究预测,如果全球平均气温较工业化前水平升高2摄氏度,每年将额

外经历至少30天热应激的0至9岁儿童人数将增至约6.5亿。此外,气候变化导致的湿热增幅在热带地区和较贫困国家尤为明显。

公报说,儿童对极端高温尤其敏感,其体温上升速度快于成年人,出汗散热能力也相对较弱。身体过热可能导致脱水、中暑等严重健康问题,长期高温暴露还可能影响学习能力和认知发展。

研究人员指出,如果不采取行动,儿童面临的损害还将继续加剧,其中发展中国家儿童尤为脆弱。他们呼吁加强对脆弱地区和人群应对气候变化的支持。

公告栏

6631203 6631205

■房产证、土地证、营运证、营业执照(个)800元 ■公章、财务章、各类印章(个)、支票、发票、收据(张)、税务登记证、资格证、代碼证、渔船用各类证件(个)及其它各类证件400元 ■身份证、工作证、户口本、出生证明、下岗证(个)150元 ■残疾证、学生证、失业证(个)20元



遗失声明

●栖霞市名郡博苑幼儿园不慎遗失银行开户许可证,开户行:栖霞中银富登村镇银行股份有限公司,核准号:J4563001256301,账号:02090130000009474,声明作废。

●烟台市宏观经济学会(统一社会信用代码:51370600MJE3018289)不慎遗失银行开户许可证,开户银行:交通银行股份有限公司烟台胜利路支行,账号:376001006010149000305,开户许可证核准号:J4560001811702,特此声明作废。