

警惕人工智能时代的“智能体风险”

【热点】

一群证券交易机器人通过高频买卖合约在纳斯达克等证券交易所短暂地抹去了1万亿美元价值，世界卫生组织使用的聊天机器人提供了过时的药品审核信息，美国一位资深律师没能判断出自己向法院提供的历史案例文书竟然均由ChatGPT凭空捏造……这些真实发生的案例表明，智能体带来的安全隐患不容忽视。

智能体进入批量化生产时代

智能体是人工智能(AI)领域中的一个重要概念，是指能够自主感知环境、做出决策并执行行动的智能实体。它可以是一个程序、一个系统或是一个机器人。

智能体的核心是人工智能算法，包括机器学习、深度学习、强化学习、神经网络等技术。通过这些算法，智能体可以从大量数据中学习并改进自身的性能，不断优化自己的决策和行为。智能体还可以根据环境变化做出灵活的调整，适应不同的场景和任务。

学界认为，智能体一般具有以下三大特征：

第一，可根据目标独立采取行动，即自主决策。智能体可以被赋予一个高

级别甚至模糊的目标，并独立采取行动实现该目标。

第二，可与外部世界互动，自如地使用不同的软件工具。比如基于GPT-4的智能体AutoGPT，可以自主地在网络上搜索相关信息，并根据用户的需求自动编写代码和管理业务。

第三，可无限期地运行。美国哈佛大学法学院教授乔纳森·齐特雷恩近期在美国《大西洋》杂志发表的《是时候控制AI智能体》一文指出，智能体允许人类操作“设置后便不再操心”，还有专家认为，智能体具备可进化性，能够在工作进程中通过反馈逐步自我优化，比如学习新技能和优化技能组合。

以GPT为代表的大语言模型(LLM)的出现，标志着智能体进入批量化生产时代。此前，智能体需靠专业的计算机科学人员历经多轮研发测试，现在依靠大语言模型就可迅速将特定目标转化为程序代码，生成各式各样的智能体。而兼具文字、图片、视频生成和理解能力的多模态大模型，也为智能体的发展创造了有利条件，使它们可以利用计算机视觉“看见”虚拟或现实的三维世界，这对于人工智能非玩家角色和机器人研发都尤为重要。

风险值得警惕

智能体可以自主决策，又能通过

与环境交互施加对物理世界影响，一旦失控将给人类社会带来极大威胁。哈佛大学齐特雷恩认为，这种不仅能与人类交谈，还能在现实世界中行动的AI的常规化，是“数字与模拟、比特与原子之间跨越血脑屏障的一步”，应当引起警觉。

智能体的运行逻辑可能使其在实现特定目标过程中出现有害偏差。齐特雷恩认为，在一些情况下，智能体可能只捕捉到目标的字面意思，没有理解目标的实质意思，从而在响应某些激励或优化某些目标时出现异常行为。比如，一个让机器人“帮助我应付无聊的课”的学生可能无意中生成了一个炸弹威胁电话，因为AI试图增添一些刺激。AI大语言模型本身具备的“黑箱”和“幻觉”问题也会增加出现异常的概率。

智能体还可指挥人在现实世界中的行动。美国加利福尼亚大学伯克利分校、加拿大蒙特利尔大学等机构专家近期在美国《科学》杂志发表《管理高级人工智能体》一文称，限制强大智能体对其环境施加的影响是极其困难的。例如，智能体可以说服或付钱给不知情的人类参与者，让他们代表自己执行重要行动。齐特雷恩认为，一个智能体可能会通过在社交网站上发布有偿招募令来引诱一个人参

与现实的敲诈案，这种操作还可在数百或数千个城镇中同时实施。

由于目前并无有效的智能体退出机制，一些智能体被创造后可能无法被关闭。这些无法被停用的智能体，最终可能会在一个与最初启动它们时完全不同的环境中运行，彻底背离其最初用途。智能体也可能以不可预见的方式相互作用，造成意外事故。

已有“狡诈”的智能体成功规避了现有的安全措施。相关专家指出，如果一个智能体足够先进，它能够识别出自己正在接受测试。目前已发现一些智能体能够识别安全测试并暂停不当行为，这将导致识别对人类危险算法的测试系统失效。

专家认为，人类目前需尽快从智能体开发应用到部署后的持续监管等全链条着手，规范智能体行为，并改进现有互联网标准，从而更好地预防智能体失控。应根据智能体的功能用途、潜在风险和使用时限进行分类管理。识别出高风险智能体，对其进行更加严格和审慎的监管。还可参考核监管，对生产具有危险能力的智能体所需的资源进行控制，如超过一定计算阈值的AI模型、芯片或数据中心。此外，由于智能体的风险是全球性的，开展相关监管国际合作也尤为重要。

新华社北京7月16日电

中国海军“和平方舟”号医院船 抵达坦桑尼亚访问并开展医疗服务

新华社达累斯萨拉姆7月16日电 当地时间16日，执行“和谐使命-2024”任务的中国海军“和平方舟”号医院船抵达达累斯萨拉姆港，开始对任务第二站坦桑尼亚进行为期7天的友好访问并提供人道主义医疗服务。这是“和平方舟”第三次访问坦桑尼亚。

上午10时许，“和平方舟”号医院船缓缓抵靠达累斯萨拉姆港。坦方在码头举行了隆重的传统欢迎仪式。坦军医疗服务部部长姆瓦米少将及部分军政官员、中国驻坦桑尼亚大使陈明健等使馆工作人员、当地民众、华侨华人和中资机构代表等在码头迎接。坦方传统舞蹈、中方舞龙等表演接连进行，现场气氛热烈。

访问期间，“和平方舟”号医院船将在主平台提供诊疗服务，同时还派出10余支医疗分队前往多个医院、社区、学校进行医疗巡诊，并与坦方人员开展联合诊疗、学术交流、文化联谊等活动。此外，“和谐使命-2024”任务指挥官拜拜会坦军政官员，参观坦军基地、医疗机构，部分官兵代表还将参观坦赞铁路、瞻仰援坦中国专家公墓。

“和平方舟”号医院船曾于2010年、2017年访问坦桑尼亚并提供医疗服务。今年是中国与坦桑尼亚建交60周年，“和平方舟”的再次到来将为进一步巩固中坦传统友谊、深化交流合作作出积极贡献。

禽流感疫情持续在美蔓延 美疾控中心派专家组协助应对

新华社北京7月16日电 美国农业部、美国疾病控制和预防中心近日表示，禽流感疫情持续在美蔓延，出现奶牛和人类感染禽流感病毒的病例。尽管如此，美疾控中心表示，目前禽流感对公众健康构成的风险仍较低。

美国农业部15日确认，俄克拉荷马州奶牛样本中发现禽流感病毒，该州成为美国第13个在奶牛中发现禽流感病毒的州。

俄克拉荷马州农业、食品和林业部门发言人李·本森说，今年4月，俄克拉荷马州一家奶牛场怀疑其奶牛可能感染禽流感病毒，遂采集样本并于近期送到美农业部检测。

美农业部在其网站公布，俄克拉荷马州有2个牛群报告禽流感病毒阳性。本森表示，这两个确诊阳性的牛群属于同一个奶牛场，目前奶牛已经完全康复。俄克拉荷马州尚未收到其他牛群感染禽流感病毒的消息。

俄克拉荷马州官方表示，该州已经为奶牛场工人提供了防护装备，并要求奶牛场针对禽流感病毒采取安全防护措施，但目前尚无针对奶牛的强制性检测措施。美国今年3月首次在奶牛中检测到禽流感病毒。此后，美国已在150多个奶牛群中检测到禽流感病毒。

科罗拉多州卫生部门14日表示，该州4名家禽养殖场工人确诊感染H5N1型高致病性禽流感病毒，另有1

例疑似病例，样本已送到美疾控中心检测。这些感染者症状轻微，包括结膜炎和发烧、咳嗽、喉咙痛等，均未住院治疗。据介绍，这些工人在科罗拉多州东北部一家暴发禽流感疫情的蛋鸡养殖场负责屠宰家禽。

美疾控中心14日发表声明称，该机构已派遣一支由流行病学家、临床医生、兽医和工业卫生学家等组成的专家组，前往科罗拉多州协助应对禽流感疫情。美疾控中心表示，目前科罗拉多州其他地区尚未出现禽流感病毒活跃度异常上升的情况。同时，针对病毒的基因组测序也在进行中，该机构将关注禽流感病毒可能出现的突变，这将影响疫情风险评估结果。

近几个月来，H5N1型高致病性禽流感病毒一直在全球野生鸟类中传播，同时已出现感染家禽和哺乳动物的情况。美疾控中心表示，虽然禽流感对公众健康构成的风险仍较低，但人群如果接触受感染或潜在受感染的动物，那么受感染风险将更大，因此建议采取相应预防措施。

美国今年已报告多例与奶牛感染禽流感病毒相关的人感染病例。英国《自然》杂志网站今年5月曾刊发文章说，一些研究人员表示，美国在应对奶牛中出现的禽流感疫情时存在数据收集和报告不足的情况，这不利于评估禽流感暴发规模，也会阻碍防止病毒进一步传播的努力。

上半年全国铁路建设持续推进

新华社北京7月16日电 盛夏时节，江苏扬州宁扬城际铁路项目建设现场机械轰鸣，100多名中铁十五局建设者在各个工地上穿梭忙碌，全力冲刺年度施工目标。宁扬城际是扬州市首条城市轨道交通线路，建成后 will 加快实现南京和扬州两市之间互联互通，促进沿线经济发展。

今年以来，全国多地铁路项目建设持续推进，取得新进展。记者从中国国家铁路集团有限公司获悉，2024年上半年，全国铁路完成固定资产投资3373亿元，同比增长10.6%，创历史新高。

上半年，铁路部门有序推进重点项目联调联试、试运行，高质量完成工程验收、安全评估、运营准备等工作，确保一批新线如期建成通车，累计新开通线路979.6公里，进一步完善了区域路网布局。其中，池州至黄山高铁开通运营后，为沿线旅游经济发展注入了新动能；兰州至张掖高铁中川机场至武威东段开通运营后，丝路重镇武威市接入全国高铁网。

与此同时，铁路部门统筹推进资源，加强施工组织和科技攻关，一批重点在建项目控制性工程取得新进展。

在辽宁，沈白高铁建设加快推进。中铁十六局项目负责人徐利表示，建设者正全面推进全线控制性工程后安隧道施工，后安隧道全长9125米，已顺利贯通进入无砟轨道施工阶段；在浙江，由中铁十一局承建的衢丽铁路衢江特大桥首根桩基顺利开钻，衢丽铁路西起衢州市连接沪昆铁路、九景衢铁路，东至丽水市贯通金温铁路，其中衢江特大桥全长超过10公里，是全线控制性工程；在安徽，由中铁十四局参建的沪宁合高铁控制性工程跨马厂水库特大桥开始架梁施工，为有序推进后续建设创造了良好条件。

“今年以来，铁路部门聚焦服务国家重大战略和区域经济社会发展，充分发挥铁路投资产业链长、辐射面广、拉动效应强的作用，优质高效推进铁路工程建设，现代化铁路基础设施体系加快构建，我国铁路网整体功能和效益进一步提升。”国铁集团有关部门负责人表示。

香港旅发局：今年上半年访港旅客量增加逾六成

据新华社香港7月16日电 香港旅游发展局16日公布数据，6月初步访港旅客数字为313万人次，较去年同期上升14%。今年1月至6月共有访港旅客约2100万人次，同比上升64%，平均每日旅客量为11.6万人次。

数据显示，今年6月，内地访港旅客约234万人次，较去年同期增加约8.6%；非内地访港旅客约79万人次，同比增加约33.6%。今年上半年，内地旅客访港约1615万人次，按年增加约6成；非内地旅客约500万人次，按年增长约8成。整体访港旅客中，过夜旅客占50%，平均留港3.2晚。

香港旅发局表示，访港旅客对香港普遍感到满意，再次访港意愿及推荐度均达到94%。香港旅发局的最新离港旅客调查问卷显示，今年前5个月访港过夜旅客到访的热门景点包括尖沙咀、庙街、太平山山顶等。



上海：在自然博物馆探秘水晶王国

7月16日，两名小朋友在纪念品商店内一组水晶摆件前驻足观看。当日，“地球宝藏之水晶传奇”特展在上海自然博物馆开幕。本次展览汇聚产自全球各地的90余件珍稀水晶藏品和宝石艺术品，将自然科普、数字媒体和艺术装置相融合，从宝石学性质、地质科学原理、文化艺术等角度为参观者呈现水晶的多维魅力。

以色列再袭加沙 援助机构警告进入加沙日益艰难

继7月13日袭击加沙地带南部导致大量人员伤亡后，以色列军方15日再次袭击加沙地带中部和南部。同日，十多家非政府组织警告，人道主义援助机构进入加沙地带日益艰难，指责以军阻碍援助物资运入。

【援助受阻】

乐施会、无国界医生组织、救助儿童会和挪威难民事务委员会等13家非政府组织15日在联合声明中说，“以色列系统性阻碍援助，并持续袭击援助行动”，以方仅为115项援助行动中的53项提供协助，不到总数的一半。声明谴责以军对巴勒斯坦伊斯兰抵抗运动(哈马斯)实施“围困战术”，加沙地带240万人口中大多数身处的以方所谓“人道主义区”已经变成“活跃战区”，“极不安全”。

声明还谴责以军近日轰炸多所用作避难场所的联合国学校。据统计，过去9天，至少6所学校遭到袭击。“非政府组织继续面临以色列持续军事行动所造成的阻碍。与此同时，最近发生的事情正加剧人道主义灾难。”

以军5月初对加沙地带南部城市拉法发起地面进攻后，拉法口岸处于关闭状态，援助物资运送“完全停止”。在加沙东南部凯雷姆沙洛姆口岸，情况同样不容乐观。

声明说，由于安全状况恶化，数吨“绝对必要”的援助物资滞留在加沙地带南部口岸。1500多辆载有药品、急救箱和基本必需品的人道主义援助卡车滞留埃及城市阿里什。与此同时，加沙地带北部的援助物资运送也“非

常有限”。乐施会说，他们花了5周才把1600个食品包从约旦运至加沙地带，而这“原本不应超过6小时”。面对多方谴责，以色列予以否认，并辩称仍有不少援助卡车进入加沙，以方还继续把物资运输受阻归咎于联合国援助机构。

【美方关切】

美国国务院发言人马修·米勒15日说，美国务卿安东尼·布林肯当天在华盛顿会见了以色列战略事务部长罗恩·德尔默和国家安全顾问察希·哈内格比，美方对以军近日袭击致大量加沙平民伤亡表示“严重关切”。双方还就加沙停火、人道主义援助和战后计划等议题展开磋商。

米勒说，以军在加沙地带造成的平民伤亡人数仍然多得“令人无法接受”，布林肯敦促以方采取更多措施以减少平民伤亡。

以军13日在加沙南部马瓦西地区发起针对哈马斯高级指挥官的清除行动，导致大量人员伤亡。加沙地带卫生部门说，以军袭击的是一片先前由其划定的“安全区”，导致至少90人死亡、约300人受伤，伤亡人数为几周来最多。

以军称此次袭击行动的目标是哈马斯军事指挥官穆罕默德·戴夫和另一名高级指挥官。两人据信参与策划去年10月7日哈马斯对以色列境内军民目标的突袭。米勒15日说，两名以方官员在会谈时告诉布林肯，尚不确定戴夫是否在以军袭击中身亡。哈马斯否认戴夫遇害身亡。

多家媒体分析，以军此次袭击行动给加沙停火谈判再次蒙上阴影。

【持续攻击】

以军近期在加沙地带攻势不减。在加沙南部，以军15日炸毁拉法西部和中部多座房屋，并在拉法东部造成10人死亡。在加沙中部，以军继续对布赖杰难民营和迈加齐难民营发动空袭和炮击。当地卫生官员说，5名巴勒斯坦人在以军对迈加齐难民营一处房屋的空袭中丧生。此外，以军还袭击了加沙中部努赛赖特难民营一处房屋，致11人死亡。

以方称，以军空袭加沙地带数十个军事目标，在拉法和加沙中部地区打死巴勒斯坦武装人员。

在加沙中部城市代尔拜拉赫，目前约70万人在这里避难。不过，当地政府15日发表一份紧急声明说，由于燃料耗尽，无法再为民众提供饮用水。

代尔拜拉赫居民瓦利德·萨比特告诉路透社记者，他家的房子在以军15日袭击中被炸毁，母亲等数名家人不幸遇难。救援人员和邻居们在废墟中搜寻幸存者。

联合国近东巴勒斯坦难民救济和工程处近日援引联合国环境规划署一份报告说，战火导致加沙地带产生约4000吨废墟，其中可能包含未爆炸的弹药和有害物质，“恐对加沙地带民众构成致命威胁”。清理废墟将耗时15年，需投入至少100辆卡车，耗资将超过5亿美元。

新华社专特稿



万人星空演唱会“点亮”鸣沙月泉

7月15日晚，游客在敦煌鸣沙山月牙泉景区参加万人星空演唱会。连日来，甘肃省敦煌市鸣沙山月牙泉景区2024年万人星空演唱会持续火爆，吸引了大量游客前来体验。每当夜幕降临，演唱会启幕，游客聚集在沙坡上跟随音乐尽情歌唱，同时打开手机闪光灯，与月牙泉夜景交相辉映。

新华社发

公告

根据国家相关法律法规及学校相关文件规定，请离校人员杜莉莉于2024年7月18日前来学校人事处补办离职手续。逾期未办理，学校将按相关规定处理。

江苏大学人事处电话：0511-88780061。

江苏大学人事处
2024年7月17日

公告

博锐拓沃精密工具(江苏)有限公司报审的年产200万件精密刀具及20万件配套刀具项目建设工程设计方案总平面图批前公示，我局将于2024年7月16日至2024年7月25日在项目现场和自然资源和规划局网站(<http://zrzy.jiangsu.gov.cn/zj>)进行批前公示，详情可在网站首页“最新信息公开”板块进行查询。

镇江市自然资源和规划局
2024年7月16日

遗失

欢迎刊登
遗失、声明、公告、特约刊登等各类信息
电话：85010151
地址：市山东路4号镇江日报社二楼
广告中心营业厅

遗失京口谏壁粮山二组朱继飞的房屋拆迁补偿安置协议，声明作废。
遗失朱家庄王贻文的房屋拆迁补偿安置协议，声明作废。
遗失京口谏壁粮山下虞范志高的房屋拆迁补偿安置协议，声明作废。