

## 警惕人工智能时代的 智能体风险

新华社记者 彭茜

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小觑。

## 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质：

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员设置后便不再操心。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家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频率。

智能体还可指挥人在真实世界中的行动。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。齐特雷恩也认为，一个智能体可能会通过在网上发布有偿招募令来引诱一个人参与现实中的敲诈案，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。智能体也可能会以不可预见的方式相互作用，造成意外事故。

已有狡猾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，它就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生产到应用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)

## 守护公众健康

## 我国举办公共卫生高层次人才专门培养项目

新华社北京7月15日电（记者 顾天成）守护公众健康，构建强大公共卫生体系，离不开人才队伍建设。邀请130多位国内外知名专家学者进行理论授课、案例教学；安排30名青年学员前往国家疾控中心、中国疾控中心、北京大学公共卫生学院等机构进行研修。2024年7月14日，由中华预防医学会主办的首期公共卫生高层次人才培养项目结业暨第二期项目启动会在北京大学医学部召开。

公共卫生高层次人才培养项目是落实党中央、国务院构建强大公共卫生体系工作部署，加强疾控人才队伍建设要求的具体举措。国家疾控局副局长高福表示，项目探索了我国公共卫生高层次人才培养模式和机制，取得可喜成绩。

中华预防医学会副秘书长李全乐介绍，公共卫生高层次人才培养项目旨在通过开展疾控体系领军人才能力提升项目和公共卫生青年精英高级研修项目，为国家、省级疾控机构和高校公共卫生学院培养一批专业功底深厚、具有战略思维和国际视野、能够解决公共卫生实

际问题的疾控领军人才和公共卫生核心技术骨干。

过去1年来，首期领军人才项目采取多次短期集中授课、专题研究等培养方式，提升学员专业领导力；青年精英项目通过7个模块课程教学、线上讲座等综合培养方式，着力提升学员公共卫生核心专业技术能力。未来，项目将不断优化课程教学，加强项目科学管理、优化培训评估体系，不断提升培训质量。李全乐说。

中国工程院院士、北京大学医学部主任乔杰表示，公共卫生与预防医学领域的突破创新关乎人类的生存和发展。坚持推进以疾病为中心向以健康为中心转变，将为全球提供中国实践和中国方案。

据悉，来自国家和省级疾控机构、部分高校公共卫生学院的专家及第一期和第二期项目学员共计150多人参加了会议。下一步，培养项目将坚持源头导向、需求导向，增强实践性，强化互补性，进一步优化培训方式方法。

## 激发科学兴趣传递科学精神

## 第六季《科学公开课》即将开讲

记者从中国科学院物理研究所获悉，第六季《科学公开课》将于7月16日晚8点正式开播，全国青少年可通过多个媒体平台观看直播。这是一档为青少年打造的纯公益、全开放的科学教育课程，旨在传授科学知识，激发科学兴趣，传递科学精神。

天文望远镜能看到怎样的景象？纳米世界是什么样的？东亚人类演化过程中发生过怎样的故事？围绕一系列富有趣味性的科学话题，约20位科学家将为广大青少年生动讲述。

据介绍，第六季《科学公开课》紧贴社会生活热点，紧跟科技发展前沿，设置了数界奇趣、理蕴万象、信息

启航、化境奇缘、生态韵律、天地奇观六个课程系列，还将针对每个主题系列课程举办科学沙龙。

公众在看过程中可通过媒体平台留言反馈疑惑和问题，主讲人或相关专家将在直播课上进行解答。直播结束后，公众也可在中国科普博览网站《科学公开课》专题页面或国家中小学智慧教育平台回看课程视频。

第六季《科学公开课》由中国科学院学部工作局、教育部基础教育司主办，中国科学院物理研究所承办，7月16日起，每周二、周四、周六晚8点在多个媒体平台同步直播，活动将持续到9月12日。（记者 张泉）

## 暑运以来全国铁路发送旅客超2亿人次

新华社北京7月16日电（记者 樊曦）记者16日从中国国家铁路集团有限公司获悉，自暑运启动以来，7月1日至15日，全国铁路累计发送旅客2.11亿人次，突破2亿人次大关，同比增长6.5%，日均发送旅客1403.4万人次。

国铁集团运输部相关负责人介绍，今年暑期学生流、旅游流、探亲流等出行需求旺盛，铁路客流保持高位运行。铁路部门认真统筹客货运输和防洪安全，精心制定暑期旅客运输工作方案，加大运输能力投放，落实便民利民惠民举措，努力为旅客平安有序出行和经济平稳运行提供可靠保障。

为保障旅客安全有序出行，各地铁路部门优化出行服务，强化路地联动机制，努力提升旅客出行体验。国铁郑州局集团公司积极协调郑州公交集团等地方市政交通部门，统筹安排公共交通运输力，方便旅客出行。最后一公里；国铁济南局集团公司与国网菏泽供电公司密切协作，对菏泽东站及管内日兰高铁等高铁线路沿线供电设施开展电力安全巡查，保障暑运期间高铁电力供应稳定可靠；针对暑期儿童研学团、旅游团出行，国铁南昌局集团公司萍乡站开辟专用进站通道和专用候车区，并安排工作人员在进站口、检票口、站台等关键位置引导帮扶，确保师生安全顺利进站乘车。



## 乐享假日

7月16日，江苏省南通市崇川区幸福街道小候鸟爱心暑托班学生在一家生物科技园的菌圃园参观灵芝。

暑假期间，各地的孩子们参加丰富的活动，乐享假日。

新华社发（许丛军 摄）

桑植县潭口灌区续建配套与节水改造工程建设项目  
公众参与第二次信息公示

根据《中华人民共和国环境影响评价法》及生态环境部令第4号《环境影响评价公众参与办法》的相关规定，我公司现将桑植县潭口灌区续建配套与节水改造工程建设项目的环评评价信息进行第二次公

示，公众可即日起登录访问<https://gongshi.qsyhbgi.com/h5public-detail?id=403985>（生态环境公示网）查阅本次公示的具体内容。  
桑植县水利建设项目建设管理中心  
2024年7月17日

## 关于永定街道片区限价商品房资格户(第五批次182户)限期购买的通知

根据市政府对征拆项目“六个清零”行动的统一安排，结合张家界市城区商品房过剩、去化周期过长的现状，市政府已明确岩塔菜市场征拆项目不再就地建设限价商品房，采取保障符合限价商品房选购资格条件的被征拆户在全城限价商品房房源中进行选购的政策，现已

全城筹集限价商品房共计1215套房源，面积139725平方米。（具体房源清单详见《房源手册》）。现依据张政办发[2020]15号《张家界市中心城区限价商品房管理办法》的相关规定，现对市中心城区永定街道片区指挥部第五批次182户限价商品房资格户，限2024年8月

5日前购房（全城购房），到期未按时购房的征收户取消其限价商品房购房资格，特此通告

限期购买资格名单明细表查询：[www.zjj-yd.gov.cn](http://www.zjj-yd.gov.cn) 检索 限价商品房

张家界市永定区住房和城乡建设局  
2024年7月16日



公益广告

创建全国文明城市

建设世界一流旅游目的地