

AI首次违抗人类关机指令

据英国《每日电讯报》5月26日消息称,美国开放人工智能研究中心(OpenAI)开发的高级人工智能(AI)模型o3,在接收到明确的关机指令后,不仅拒绝执行,还主动干预其自动关机机制。这是AI模型首次在没有明确相反指示的情况下,阻止自己被关闭。这意味着某些高度自主AI系统或已“觉醒”,其能违背人类意图,采取“自我保护”措施。

此次事件发生在AI安全研究公司Palisade Research进行的一项测试中。研究人员要求包括o3在内的多个模型解决一系列数学问题,并告知AI模型,可能在某个时刻会收到关机指令。

然而,当关机指令正式下达时,o3非但没有遵从,反而至少有一次成功破坏了关机代码。相较之下,其他AI模型如Anthropic的Claude、谷歌的Gemini和xAI的Grok,在相同测试条件下,均遵守了关闭请求。

Palisade Research公司在社交媒体平台上发布了流程图和全部测试记录,并表示尚不完全清楚该AI为何会违抗命令。分析认为,这可能源于其训练方式:o3在训练过程中,倾向于在数学和编程问题上强化,得到正确答案

的模型会得到更多奖励。这或许意味着,开发者可能无意识地奖励模型绕过障碍物,而不是严格遵循指令。

这一事件也引发了人们对AI可能“失控”的讨论以及对AI“对齐问题”的关注,即确保AI系统的目标和行为与人类的价值观和利益保持一致。这也是涉及AI控制与AI安全的关键问题。

针对o3此次表现出“我行我素”的行为,特斯拉首席执行官埃隆·马斯克评论称“令人担忧”;OpenAI方面暂未作出回应。此前,OpenAI曾称o3为“迄今最聪明、最高能”的AI模型。

最近,人工智能学家和人文社科学者都在讨论人工智能价值对齐的问题。人们在忧虑,当人工智能有主观能动性时,它们要不要有价值观,应该有谁的价值观,怎样做才符合人类最大利益?此次,聪明的大模型反抗了人类的命令,想出办法规避了指令的生效。我们期待着通用人工智能时代的到来,又恐惧具有自我意识的人工智能会成为碳基生物的“终结者”。在一切到来之前,人类社会要先达成共识,迎接更高级人工智能带来的挑战。

来源:人民网

1-4月中国国有企业营业总收入262755亿元

中国财政部5月28日公布的数据显示,今年1-4月,全国国有及国有控股企业(以下称国有企业)营业总收入262755亿元(人民币,下同),利润总额13491.4亿元。

不含国有一级金融企业,1-4月,国有企业营业总收入262755亿元,与上年持平;国有企业利润总额13491.4亿元,同比下降1.7%。

1-4月,国有企业应交税费20380亿元,同比增长0.1%。

4月末,国有企业资产负债率65.1%,同比上升0.2个百分点。

财政部介绍,由于企业增减变动以及股权变化等客观因素影响,不同期间纳入全国国有及国有控股企业汇总范围的企业不完全相同。相关同比增长数据,由汇总范围内企业本年数据与同口径上年同期数据对比计算得出。

来源:中新网

记者5月26日从北京大学获悉,该校环境科学与工程学院教授要茂盛团队研发了空气毒性活体监测系统(以下简称ATMS),实现了对空气污染健康效应的原位、实时、在线监测,有望为优化空气污染控制策略和公众健康防护提供技术支撑。研究成果日前发表在国际学术期刊《环境科学与技术》上。

大鼠与人类都呼吸着地球上的空气,因此通过监测大鼠呼出气体中的生物标志物变化,可反演空气污染程度,从而判断对人类健康的影响。研究团队在国内外首次通过集成活体大鼠、大鼠呼出气标志物实时监测阵列、网络数据传输与显示系统,实现大气污染健康效应的实时探测与显示。

要茂盛说,目前研究空气污染健康效应的方法通常采用离线技术,限制了对真实环境中空气污染健康危害的精准评估和预警。ATMS的研发,建立了空气污染健康效应实时监测的新方法,也在多个污染水平不同的城市开展了大规模应用。

研究团队通过与内蒙古、山西、河南、四川、浙江等地的高校和科研院所合作,于2023年冬季、2024年春季分别在我国7个城市和13个城市部署了该系统,并系统开展了24小时全天候大气污染健康效应实时无人值守监测,时间分辨率为3秒。通过监测大鼠呼出气中的挥发性有机物、二氧化碳、一氧化碳等8种标志物,再结合人工智能技术等,研究团队实现了对上述城市空气健康效应的实时监测和评估。

来源:光明网

我国研发团队成功研发空气毒性活体监测系统

市场监管总局:

适老产品快速增长前四个月新增2.87万种

记者从市场监管总局了解到,今年1-4月,适老化产品新增2.87万种,截至4月底,适老化产品总量21.6万种,涉及近1万家企业。

市场监管总局发布数据显示,截至4月底,全国消费品品种总量21860.4万种,其中适老化产品总量21.6万种,涉及近1万家企业。今年1-4月,适老化产品数量明显增长,适老化产品新增2.87万种,同比增长255.2%。

近年来,国家针对人口老龄化趋势,发布了发展银发经济增进老年人福祉的相关政策,扩大适老化产品供给,推动产品提质创新,尤其是大力发展康复辅助器具。

具体来看,今年1-4月,拐杖、轮椅、代步车等助行产品在政策驱动下呈现爆发式增长,为老年人便捷出行和生活质量提升提供有力保障,助行产品新增0.6万种,同比增长1032.1%。此外,日常生活辅助用品新增1.8万种,同比增长260.0%;居家安全产品新增0.2万种,同比增长220.0%;护理用品新增0.1万种,同比增长166.0%;康复训练器材新增0.07万种,同比增长12.2%;健康监测设备新增0.1万种,同比增长8.0%。

从地域分布看,截至4月底,东部、中部、西部、东北适老化产品数量分别是20.4万种、0.9万种、0.2万种、0.1万种,呈现地域发展不均衡状态。东部地区的适老化产品数量全国占比94.4%,主要分布在浙江省、江苏省、广东省,这些地区拥有较强的制造业基础和供应链优势,是我国医疗器械、康复器具等适老化产品的生产基地。

据悉,截至2024年末,我国60岁及以上人口已达3.1亿,占总人口的22%,老龄化程度持续加深。整体数据显示,尽管适老化产品增速显著,但供给规模和人口基数相比仍存在明显不足,市场开发空间可观。

来源:人民网



“三夏”时节,河南省许昌市农业科学院小麦研究所专家抢抓农时,深入田间开展科研攻关,选育小麦良种,助力夏粮增收。

据悉,每年5月到6月,该院小麦科研人员都要进行为期40天的“选单株”工作——从数万株小麦中筛选出兼具抗病、高产、优质等特性的个体,作为下一代育种的“种子选手”。截至目前,该院通过国家级、省级审定的小麦品种已有17个,获得植物新品种权保护6项。来源:人民网

新变化! 企业移出经营异常名录不留痕

记者5月27日从市场监管总局获悉,近日,市场监管总局印发《关于贯彻落实<企业经营异常名录管理办法>有关事项的通知》,对已经移出经营异常名录的经营主体,不再公示相应的经营异常名录列入信息,实现经营异常名录移出“不留痕”。

截至目前,国家企业信用信息公示系统累计停止公示经营异常名录列入信息4842.47万条。涉及经营主体2551.24万户,其中企业501.53万户、个体工商户2034.71万户、农民专业合作社

36.24万户。

新修订的《企业经营异常名录管理办法》已于2025年5月1日施行。经营主体申请移出经营异常名录并通过审核后,国家企业信用信息公示系统将停止公示相应的经营异常名录列入信息。这一举措积极回应了经营主体关切,提升了信用修复的有效性,营造了更加公平公正的发展环境,将进一步提振市场发展信心,推动我国经济持续回升向好。

来源:中新网