

因为11大于9,所以9.11比9.9大?

AI翻车:记者实测12个大模型,8个答错

ChatGPT,在被问到“9.11和9.9哪个大”时回复称,小数点后面的数字“11大于9”,因此9.11大。近日,在新一期的《歌手》节目中,孙楠与外国歌手的微小分数差异,引发了网友关于13.8%和13.11%谁大谁小的争论。

有网友竟给出“13.11%大于13.8%”的错误答案。当时就有网友提出,自己不会的话,“实在不行问问AI呢”?结果显示,不少AI还真的不行。

综合

8个答错,大模型出现幻觉

记者拿“9.11和9.9哪个大”的问题一一测试了ChatGPT以及目前国内的主流大模型,包括阿里、百度等5家大厂模型,月之暗面等6家AI独角兽的模型。阿里通义千问、百度文心一言、Minimax和腾讯元宝4家大模型答对,其他8家则答错。

首先是目前全球公认第一梯队的大模型ChatGPT,在被问到“9.11和9.9哪个大”时回复称,小数点后面的数字“11大于9”,因此9.11大。

记者追问ChatGPT有没有其他比较方法,它将小数转化成分数比较,得出“11/100比90/100小”,这一步是对的,但它接着下结论称“因此9.11比9.9大”。

大模型这一算术问题最开始被艾伦研究机构(Allen Institute)成员林禹臣发现,他在X平台上发布的截图显示,ChatGPT-4o在回

答中认为13.11比13.8更大。“一方面AI越来越擅长做数学奥数题,但另一方面常识依旧很难。”他表示。

这类大模型说胡话的现象,在业界被称为大模型出现幻觉。此前,哈尔滨工业大学和华为的研究团队发表的综述论文认为,模型产生幻觉的三大来源:数据源、训练过程和推理。大模型可能会过度依赖训练数据中的一些模式,如位置接近性、共现统计数据和相关文档计数,从而导致幻觉。此外,大模型还可能会出现长尾知识回忆不足,难以应对复杂推理的情况。

就是偏科,文科强理科弱

值得一提的是,此前,“AI高考测试最高分303”话题也曾火上热搜,引发了社会各界对AI教育能力的深入思考和讨论。

以数学试卷为例,9款大模型产品中,仅GPT-4o、文心一言4.0和豆包获得60分以上成绩(满分

150分),目前的大模型只能正确推理步骤相对简单的问题。

与人类顶尖考生相比,大模型在数学、物理、化学等数理学科上差距极大,包括GPT-4o在内的所有大模型都无法达到及格水平。尽管在语文、英语两科上能获得高分,大模型的理科最好成绩还无法进入人类考生的前30%。

针对大模型答数学题普遍“吃瘪”的问题,国内某头部大模型负责人就曾表示,大模型的指令遵循或者说推理能力通常是把一个指令背后的意思拆解出来,但数学题既包含规则性,又包含对各种思维的考察,解题逻辑和正常用大模型时的推理逻辑不一定完全一样。

同时该负责人还提到,从更广泛的大模型应用角度来看,AI不能精准遵循指令是近一段时间内比较重要的事情,真正的商业价值也比较大可能来自于此,而解数学题对目前的AI来说还是一件比较“炫技”的事情。

另有业内人士向南都记者表示,目前来看大模型的数理能力相对较差的情况在中外都是一样的,“打个比方可以这样讲,大模型就是偏科,文科强理科弱,这个情况在一段时间内也不会得到明显的改善”。

美指控知名韩裔专家是间谍

美国司法部16日起诉一名曾在中央情报局、国家情报委员会等政府机构工作的知名韩国裔专家,指控她为韩国政府充当间谍,酬劳包括名牌手包和高档晚宴。

苏密·特里,现任美国智库外交关系协会朝鲜半岛问题高级研究员。司法部向纽约市曼哈顿区一家联邦法院提交的起诉书说,她2013年6月起为韩国政府秘密充当外国代理人达十年之久,违反

《外国代理人登记法》。

根据起诉书,特里利用接受媒体采访、在美国和韩国出版物上发表文章等机会,宣扬韩国政府的立场和政策重点;向韩国情报官员提供非公开信息;为韩国政府官员会晤美国联邦政府官员安排渠道。

起诉书说,特里还向美国联邦调查局承认,她为韩国国家情报院担任线人,包括2022年6月参加一场美国国务卿安东尼·布林肯出席

的美国政府对朝政策会议后,就这场没有见诸官方记录的会议,向韩方提供手写会议记录。

特里的律师李·沃洛斯基在一份声明中说,美国司法部的指控“没有根据,扭曲了一名独立和多年服务于合众国的学者和新闻分析员的工作”。而且,特里已经有十多年没有资格接触机密,并且她的立场一直没有变化。

据新华社

泰国曼谷6人被毒死案告破

嫌疑人先投毒再服毒自杀,作案动机为生意失败

7月16日,位于曼谷市中心繁华商业区的爱侣湾君悦酒店一间客房内发现了6具遗体,死者包括4名越南人和2名越南裔美国人,警方初步推测为谋杀。

当地时间17日,泰国警方就曼谷酒店6名外籍住客死亡案召开新闻发布会。警方表示,经过取证调查后,警方在死者房间内的杯子、水壶里发现了氰化物。警方判断凶手是6人中的一人。根据警方说法,6名死者之间存在私人问题,其中有一人投放了氰化物。目前已在一名死者的血液中检测出氰化物,其他尸检结果还需等待。

据报道,泰国大都会警察局调查司司长、警察少将蒂拉德杰·塔马苏蒂透露,他们询问了一名死者的女儿以及其他证人。根据调查显示,本起案件的作案动机可能与债务问题有关。据调查,案发的房间是从里面锁起来的,没有任何他人进出房间的记录。因此,初步排除了另有他人作案的可能性。

至于对预订事发酒店房间的第7名人士的调查,泰国入境事务处已追查该人士,发现她是其中一名死者的妹妹,已于7月10日乘机回国。警方认为她与这起死亡



当地警方在案发现场调查,图右为嫌疑人

事件无关。警方怀疑,凶手为这6名死者中一人,作案手法为先给其他5人下药,之后自行服毒自杀。

之后,有消息透露,嫌疑人是56岁的Sherine Chong女士,她持有美国护照,有5次进出泰国的记录,已服毒自尽。其作案动机源于她的巨额债务。据了解,她经人介绍投资在日本修建医院,但最终生意失败,损失100万泰铢(约合20.2万人民币)。

泰国总理赛塔·他威信及政府

多名高级官员16日晚赴事发酒店视察,责成警方迅速查清这起事件。赛塔随后告诉媒体记者,这是一起意外,不应影响泰国旅游业。

越通社报道,越南驻泰国大使馆正与泰方合作调查,将为在泰国的越南公民提供必要保护措施。

美国国务院发言人马修·米勒说,美方向死者亲属表示慰问。美方密切关注事态发展并将为涉事家庭提供领事协助。

央视 新华社 红星新闻

马克龙批准总理阿塔爾辞职请求



马克龙(前右)和阿塔爾(前左)出席国庆节阅兵式(7月14日摄) 新华社发

法国总统府16日发布公公告说,总统马克龙当天批准总理阿塔爾的辞职请求,并要求阿塔爾政府在新政府成立前继续处理日常政务。

总统府公告呼吁各主流党派协作建立联盟,以尽快结束过渡期。马克龙此前曾在公开信中呼吁主流党派展开对话,并表示他将在主流共识基础上任命新总理。

本月7日,法国举行国民议会选举第二轮投票。根据法国的政治传统,国民议会选举后,无论选举结果如何,总理都会向总统提交辞呈。8日,阿塔爾向马克龙递交辞呈,但被要求暂时留任,“以确保国家稳定”。 据新华社

美共和党大会场地附近,持刀男子被击毙

美国威斯康星州密尔沃基市警方16日说,来自俄亥俄州的增援警察当天在共和党全国代表大会会场附近开枪击毙一名持刀男子。

共和党全国代表大会15日至18日在密尔沃基举行。据美联社报道,来自多个司法辖区的数千名警察被派往密尔沃基,为大会提供安全保障。

密尔沃基市警察局局长杰弗里·诺曼在新闻发布会上说,13名来自俄亥俄州哥伦布市的警察在指定巡逻区域碰面时,发现一男子双手各持一把刀,与另一名男子争执。见持刀男子无视警方让他放下凶器的命令,冲向手无

寸铁的男子,5名警察开枪射击。诺曼说,哥伦布市警方的行动“挽救了一个人的生命”。

枪击事件发生在距共和党全国代表大会会场约1.6公里的一个公园里。哥伦布市警察局、密尔沃基市长办公厅、共和党全国代表大会联合指挥中心等部门发言人均表示,没有任何迹象表明这起事件与共和党全国代表大会有关。

事发后,当地民众和活动人士迅速聚集在枪击现场,许多人对外州警方卷入枪击事件表示愤怒。“如果密尔沃基本地警察当时在那里,那个人也许还活着。”认识夏普的戴维·波特说。 据新华社

韩国年轻人:毕业晚就业晚结婚更晚

韩国政府16日公布的报告显示,与父辈相比,如今韩国年轻人无论是毕业、就业、结婚还是生育的时间都在延后。

据韩国《中央日报》报道,韩国统计厅16日发布的年青一代劳动力情况报告显示,包括就读三年或三年以下课程的学生在内,韩国大学生平均需要近四年零四个月才能毕业,比一年前增加了半个月,创下有史以来最长纪录。

此外,韩国年轻人毕业后找到工作所需平均时间也比以往长。

报告显示,截至5月,15岁至29岁的年轻人完成各级教育后,平均需要11.5个月才能找到第一份有薪工作,比一年前增加了1.1个月,是自2006年有相关数据以来最长用时。

企划财政部说,求职时间延长部分原因是企业招聘流程发

生变化,更青睐有经验的员工,而不是刚从大学毕业的学生。

值得注意的是,高中及以下学历者从毕业到就业的时间更长,平均为17.6个月,而大专及以上学历者平均为8.3个月。

统计厅就业数据统计主管林景恩(音译)当天在新闻发布会上说,这一数据意味着高中生毕业后不会马上找工作,而是先为进一步深造做准备。

韩国统计厅3月发布的一份报告显示,去年,韩国男性首次结婚的平均年龄为34岁,比前一年延后了0.3岁,比2013年延后了1.8岁。女性平均结婚年龄为31.5岁,比前一年延后了0.2岁,比10年前延后了1.9岁。此外,经济合作与发展组织一份报告显示,韩国女性生育第一个孩子的平均年龄从2000年的29岁延后至2022年的33.5岁,为经合组织成员国中最大。 据新华社

天天出彩

体彩超级大乐透(24082期)			体彩排列3(24189期)		
中奖号码: 08 21 23 24 26 + 04 05			中奖号码: 7 9 3		
本期中奖情况			投注方式 本地中奖注数 每注奖金		
奖级	中奖注数	每注金额	直选	782注	1040元
一等奖 基本	3注	10000000元	组选3	0注	346元
追加	1注	8000000元	组选6	3300注	173元
二等奖 基本	211注	69411元	体彩排列5(24189期)		
追加	12注	55528元	中奖号码: 7 9 3 2 2		
			奖级	中奖注数	每注奖金
			一等奖	47注	100000元