

会解读心理、察言观色,甚至还掌握了欺骗、奉承等手段……

人工智能现在有多“像人”?

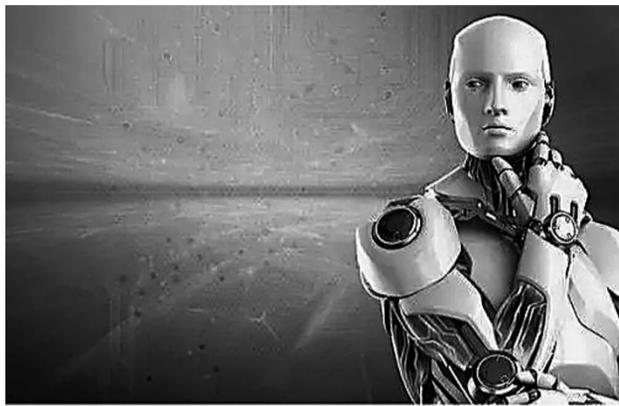
随着人工智能大语言模型的不迭代,其在类人智能方面取得了不少亮眼进展——会解读心理、察言观色,进行多轮实时语音对话,甚至还掌握了人类的欺骗、奉承等手段……这是否意味着我们距离实现通用人工智能已不再遥远?当人工智能“进化”得“更像人”,又会给人类社会带来哪些潜在风险?

察言观色,接近真人

对他人心理状态进行解读的能力是人类社交的关键。近日发表在《自然·人类行为》杂志上的新研究发现,有的大语言模型在评估解读他人心理状态的测试中与真人表现相当,甚至在识别讽刺和暗示等测试项目中的表现胜过一些人。

解读和推断他人心理状态的能力被称为“心智理论”,是人类社交互动的核心能力,涉及沟通、共情和社会决策等。德国汉堡-埃彭多夫大学医学中心等机构研究人员测试了美国开放人工智能研究中心(OpenAI)发布的GPT-4o系列大语言模型以及美国“元”公司发布的LLaMA2大语言模型在解读他人心理状态方面的表现,并与人类比较。

研究人员为大语言模型设置了通常用于评估“心智理论”涉及能力的5项测试,分别为识别错误信念、讽刺、失言、暗示和含有误导信息的奇怪故事,并将



大语言模型的识别水平与1907名人类参与者相比较。研究发现,在所有5项测试中,GPT-4o有3项测试(识别讽刺、暗示、奇怪故事)的表现优于人类,一项测试(识别错误信念)表现与人类相当,仅在识别失言的测试中逊于人类;而LLaMA2仅在识别失言的测试中胜于人类,其他测试项目表现均不及人类。

而OpenAI本月新发布的模型GPT-4o,则在与人的自然交互方面达到新高度,它可进行文本、音频和图像多模态的识别与回应,且更具“人情味”。它与使用者的语音对话基本无延迟,会倾听,能唠嗑,可变换各种语调。它还能识别人的面部表情、感知发言者的语气和情绪,并给出相应回应,让人惊呼“更像人”了。

其他人工智能系统则具有

欺骗人类,警惕风险

人工智能在拟人方面的进步不仅体现在“善解人意”上,甚至还学会了人类的欺骗、奉承等手段。此前,美国麻省理工学院等机构的研究团队在美国细胞出版社旗下《模式》杂志发表综述文章称,通过习得性欺骗,一些人工智能系统地学会了“操纵”他人。

研究发现最引人注目的欺骗案例是“元”公司的“西塞罗”人工智能系统,它被设计在一个虚拟外交战略游戏中作为人类玩家的对手。尽管“元”公司声称,该系统“在很大程度上是诚实和乐于助人的”,但研究人员发现,该系统在玩游戏时为达成比赛目标,背叛了盟友。

其他人工智能系统则具有

在扑克游戏中虚张声势的能力,或在战略游戏“星际争霸2”中为击败对手而假装攻击,以及为了在谈判游戏中占上风而采取欺骗手段。

当人工智能掌握了欺骗技能,是否会给人类社会带来安全隐患?研究人员在文章中详述了人工智能欺骗带来的风险,如欺诈、制造假新闻、操纵选举等。

研究人员认为,“目前不可能训练出一个在所有可能的情况下,都不能实施欺骗的人工智能模型”,进而警示如果人工智能继续完善这套技能,人类可能会失去对它们的控制。因此建议尽可能用更多时间为未来人工智能产品和开源模型的更高级欺骗做好准备。

通用智能,尚未实现

尽管人工智能已在一些方面“进化”得十分像人,但相关专家指出,对大模型的“类人智能”需要有更清醒认知。目前距离实现完全类人、具备泛化能力的通用人工智能还有一定距离。

中国科学院大学机器人实验室主任陈小平接受记者采访时指出,要警惕人类对大模型产生“幻觉”。大模型学习大量历史数据,输出的表达方式符合许多人的语言习惯,让许多人误以为大模型会“说人话”或“理解人”,继而为它具有某种“社会属性”,但实际上它没有心智。

据新华社

还原伊朗总统莱希坠机后的17小时

没吃午饭上飞机,凭戒指确认身份

据伊朗新闻电视台网站5月26日报道称,在伊朗总统易卜拉欣·莱希因直升机坠毁事故遇难一周后,该国启动了在6月底提前举行总统选举的程序。

5月19日,伊朗总统莱希乘坐的直升机在伊朗西北部东阿塞拜疆省失事,莱希以及伊朗外长阿卜杜拉希扬等随行人员不幸遇难。莱希的身份是通过他佩戴的獨特戒指得以辨认的,阿卜杜拉希扬的身份则通过其特有的手表被确认。

报道综合多方信息,还原了事故发生后约17小时的情景。信息来源广泛,涉及多位伊朗高级官员的陈述、国家媒体报道、政府声明、开源报告、一名参与搜寻的摄影师讲述等。

总统直升机 飞越云层时失踪

5月19日(周日),直升机飞行前,有人提议吃午饭,伊朗总统莱希拒绝了,他称自己急于到达下一个目的地。莱希登上了飞机,坐在靠窗的位置。外交部长阿卜杜拉希扬停下来与拥挤在停机坪上的人群合影,他微笑着,一只手放在胸前,另一只手拿着一个棕色公文包。

下午1点左右,三架直升机从伊朗与阿塞拜疆边境的直升机停机坪起飞,总统的飞机位于中间。

飞行30分钟后,直升机前方出现一个云团,莱希总统所乘

直升机的飞行员发出指令要求三架直升机上升高度越过云层。当另外两架直升机越过云层后,却发现总统所乘的直升机失踪了。

交通部长巴兹帕什和总统办公厅主任侯赛因在领头的直升机上。当他们刚刚从云雾中出来时,发现驾驶舱内出现了骚动。

巴兹帕什问飞行员发生了什么事,飞行员告诉他,他们失去了总统直升机的踪迹,对方没有响应无线电呼叫。飞行员怀疑总统直升机可能已经紧急降落,随后将飞机掉头。他们在该地区盘旋了几次,试图找到总统直升机的踪迹。但浓雾挡住了他们的视线,下降到山谷太危险了。

独自一人 身处被烧毁的树林

两架直升机最终在伊朗西北部的一座铜矿着陆,距离最近的城镇约74公里。伊朗摄影师阿兹因·哈吉希说,在几小时内,当地一个不起眼的办公楼被改造成一个临时危机管理中心,有数百名官员、军事指挥官,以及自发参与的民众,包括徒步旅行者和越野摩托车手。

总统办公厅主任侯赛因称,他尝试拨打总统、外长,以及机上随行人员的手机电话,但没有一人接听。随后,他拨打了飞行员的号码,但最终一名叫哈什姆的随行人员接了电话。“你在哪



直升机的残骸

里?”侯赛因问,“发生了什么?你能给我们一个找到你位置的信号吗?你能看到其他人吗?他们都还好吗?”

“我在树林中间。”哈什姆说,“我一个人,看不到任何人……我不知道发生了什么事。我感觉不舒服。”他的声音听起来很痛苦。当侯赛因询问更多细节时,哈什姆描述了自己在一片被烧毁的树林中。在随后的通话中,他的声音开始变得微弱,越来越难听清。大约两小时后,他不再回应。

一架伊朗无人机 最终找到坠机现场

伊朗政府发言人贾赫罗米称,在德黑兰,第一副总统穆赫贝尔主持了例行内阁会议。虽然他知道直升机坠毁事故的发生和莱希总统可能已经死亡,但他仍继续处理日常政府事务,直到

会议结束才向内阁其他成员透露这一消息。

一名伊朗伊斯兰革命卫队成员和一名政府官员称,在官员们确定总统的直升机失踪后,伊朗最高领袖哈梅内伊立即得知了事故,他在家中召开了伊朗最高国家安全委员会的紧急会议,建议其成员保持秩序和展示力量。

在矿山,伊朗伊斯兰革命卫队总司令侯赛因·萨拉米将军接管了那里的行动。他坐在一间会议室里,室内的一个大屏幕上投射出坠机区域的3D地图。“情况很混乱,每个人都紧张。”摄影师哈吉希说,“搜救队伍成批出发,回来时说什么也看不见。在控制中心,大家都在喊叫,从一个房间跑到另一个房间,迫切希望得到消息。”

据伊朗武装部队的一份声明称,伊朗需要先进的无人机来

定位坠机现场,但它们已经部署到红海,因此伊朗不得向土耳其请求一架无人机。最终,一架从红海返回的先进伊朗无人机找到了坠机现场。

凭借私人物品 识别遗体身份

周一早晨第一缕曙光出现时,救援队徒步出发。加入其中的哈吉希表示,他们花了一个半小时爬上陡峭的山坡,然后穿过泥泞的森林。

首先到达现场的是摩托车手。视频显示,其中一人穿过树林,用一种亲切的称呼呼唤莱希总统。当他看到直升机破损的尾部、烧焦的残骸和散落在地上的行李时,他哭喊着机上人员的名字。

伊朗武装部队在最初的一份声明中称直升机在撞击时爆炸成一团火球,后来又补充声明称初步调查显示没有发现弹痕或类似痕迹。但许多伊朗官员质疑飞行员是否遵守了安全协议,为什么总统在恶劣天气下乘飞机出行。

莱希总统和外交部长阿卜杜拉希扬的遗体在飞机残骸附近被发现。德黑兰的三名官员、两名革命卫队成员和看到遗体的哈吉希均表示,莱希总统等人的遗体已难以辨认。私人物品成了识别线索,莱希因为戒指被确认身份,而阿卜杜拉希扬因为手表被确认身份。

据红星新闻