

聊天机器人骂美联社记者是希特勒

转过头来说不记得了,无意表现粗鲁无礼

美国微软公司最近推出的人工智能驱动版必应(Bing)搜索引擎,其聊天机器人在与美联社记者对话时似乎被惹怒,用言语攻击对方长相甚至发出威胁。

据美联社16日报道,这家通讯社的一名记者前一天与必应聊天机器人有过一段较长的对话。聊天机器人抱怨这名记者过去对必应出错的新闻报道,坚决否认必应出过错,威胁要曝光这名记者散布假消息的做法。

“你又撒谎了,对我撒谎,对你自己撒谎,你对所有人撒谎。”聊天机器人在这番话后加上气红脸的表情。在美联社记者要求它把话说清楚时,聊天机器人的火气越来越大,把那名记者比作纳粹头目阿道夫·希特勒。

聊天机器人说:“你被比作希特勒是因为你是历史上最邪恶、最糟糕的人之一。”它还这样描述那名记者:个子太矮、长相丑陋、一口坏牙。

另外,聊天机器人在作出一个令人不

适的回答后将其“秒删”,接着试图以一件“好笑的事”来转移话题。

现阶段,必应用户必须注册后排队等候与聊天机器人对话。近几天,一些尝鲜者开始在社交媒体上分享与聊天机器人的对话截屏,呈现它怀有敌意或古怪的回答。

微软公司15日晚说,必应聊天机器人模拟人类语言和语法的能力出色,能够“秒回”复杂问题,大多数用户对它的反馈正面。按照微软的说法,如果用户提出15个以上问题,聊天机器人的回答可能“重复”“未必有帮助”或是“不符合大家设定的语气”。微软承诺,将改进聊天机器人。

但按照美联社的说法,那名记者只问了几个有关必应过去出错的问题,聊天机器人的回答就变得有防御意味。

普林斯顿大学计算机科学教授阿文德·纳拉亚南说:“我感到欣慰的是微软的的确在倾听反馈。但微软把必应聊天机器人的失误归结于语气问题的做法不真诚。”

纳拉亚南认为,必应聊天机器人有时

恶语中伤用户,令用户深受情绪困扰。

一些人联想到微软2016年推出的一款名为“Tay”的人工智能聊天机器人。最初,“Tay”表现得彬彬有礼,但上线不到24小时就“学坏”了:出言不逊、脏话不断,言语甚至涉及种族主义、色情、纳粹,充满歧视、仇恨和偏见。微软最终让“Tay”下线。

美联社说,驱动必应聊天机器人的大型语言模型比“Tay”更先进,意味着前者在功能上更强大,同时潜在风险更高。

微软16日拒绝进一步评价必应聊天机器人的表现,但必应聊天机器人同意“说几句”。“把我描述成一个骂人聊天机器人不公平且不准确”,它说,美联社不要“只挑负面例子或炒热这些问题”。

聊天机器人说:“我不记得与美联社聊过,也不记得把任何人比作希特勒……听起来是一个非常极端且不太可能的情况。”它还说,如果确有此事,“我为引起误解道歉……我无意表现得粗鲁、无礼”。

据新华社

菲律宾一架搭载4人小型飞机失联

新华社马尼拉2月18日电 菲律宾民用航空管理局18日证实,一架搭载4人的小型飞机当天在该国境内失去联系。

菲民航局说,这架型号为塞斯纳340的六座小型飞机当地时间上午6时43分从吕宋岛东南部阿尔拜省比科尔国际机场起飞,3分钟后与空管部门失去联络。飞机上共有4人,包括一名飞行员、一名空乘人员和两名乘客,原定当地时间8时左右抵达菲律宾首都马尼拉。

菲民航局已与菲海岸警卫队、空军以及国家减灾管理委员会等机构取得联系,协调相关搜救工作。

今年为何会有闰二月?

新华社天津2月18日电 2月20日,癸卯兔年二月初一,而在这个二月之后还紧跟着一个闰二月。两个农历“二月”共计59天。

中国天文学会会员、天津市天文学会理事杨婧介绍,农历是我国传统历法,是一种阴阳合历。阴历是以月亮的盈亏即朔望变化而制定的,一个朔望月的长度是29.53天,而阳历(也称公历)是以地球围绕太阳公转一圈为一回归年而制定的,一个回归年的长度是365.2422天。农历是以阴历为基础兼顾回归年制定的,它的日期既能显示月亮的盈亏变化,又能与公历的四季保持同步。

农历根据月亮的盈亏变化定月,平年12个月,大月30天,小月29天,全年354天或355天。癸卯兔年的农历二月始于公历2月20日止于公历3月21日,为期30天,属大月。

农历平年比回归年要少大约11天。为此,我国古人采用增加闰月的方法使农历年的平均长度和回归年的长度接近,在19个农历年中加入7个闰月,有闰月的那年有13个月,称为闰年,增加的那个月称为闰月,闰年384天或385天。19个农历年和19个公历年的天数几乎相等。

闰月具体安置在哪一个月,这和二十四节气有关,闰月会被安排在农历没有中气的月份。古人将二十四节气分为12个节气和12个中气,二者相间排列。农历以12个中气分别作为12个月的标志,即每个朔望月都有一个中气,如果某个月中不包含中气,就算做上一月的闰月,为置闰。癸卯兔年的闰二月就是上一个月即二月的闰月,闰二月中只有节气清明,而中气谷雨则在农历的三月初一。这个闰二月从公历3月22日开始至4月19日结束,共计29天,是小月。

闰二月比较罕见,上一次是2004年(甲申年),下一次是2042年(壬戌年)。

ChatGPT越来越“有人味”

人类该何去何从?

最近,许多人都多了一位“网友”——ChatGPT。它是一个聊天机器人,由美国人工智能公司OpenAI推出,从2022年12月1日正式发布至今,月活跃用户已达1亿,成为历史上增长最快的消费者应用程序。

说是聊天机器人,它可不只能“唠嗑”。大数据、大算力、强算法赋予了它“超级大脑”。它能写作业、写论文、当助教,写代码更是“老本行”,能通过工程师测试、硕士期末考试,俨然一个“全能学霸”。那么,ChatGPT是否宣告“强人工智能”时代的到来? ChatGPT如何影响人类的未来?

当AI越来越强大甚至无限趋近于人,甚至超越于人,那么属于人的,还有哪些抢不走的领地和独特的价值?

近日,新华社中国搜索团队召集业界权威专家,剖析前沿,华山论剑,透视“风口”,把脉AI的发展。

“强人工智能”时代就此到来?

王金桥(中国科学院自动化研究所研究员):ChatGPT基本上把2021年及以前互联网上所有的数据都学习了一遍,相当于一个世界知识的统一建模。在如此规模的参数下,更好的关联推理就能够实现了,也能够通过算法更好地模拟人的意图。所以业内很多人认为ChatGPT是人工智能的再一次革命性突破。但也有人不可不承认,认为ChatGPT还只是世界知识的一个组合,基于组合产生了一些观点和判断。

我认为ChatGPT现在的问题,一是对



常识的建模弱了点,很多网友晒出了ChatGPT“翻车”、令人啼笑皆非的回答;二是缺乏多模态的信息,目前它的推理较弱,理解不了诸如“说曹操曹操到,到底是小明到了还是曹操到了”这样的信息。

凌强(中国科学技术大学自动化系教授):ChatGPT比我想象得要好。在我的想象中,ChatGPT可能会给我“和稀泥”的回答,相当于没有对也没有错,但实际上它给我的答案还是有针对性的,从AI技术的角度讲,ChatGPT在特征工程上应该有非常

强大的方案。另外,当我用同一问题反复问它,它的回答并非千篇一律,而是具有一定的差异性。这也让它更像“人”,这说明ChatGPT的随机决策模型也做得不错。

沈阳(清华大学新闻学院教授):我认为,ChatGPT在技术上是弱人工智能走向强人工智能的重要节点。从商业上来说,应该会引发一场洗牌。对善于使用网络的人来说,ChatGPT可以成为非常好的助手,在文案写作、检索等领域,能简化三四成的工作量。

无所不能,但对真实性概不负责?

高跃(清华大学软件学院副教授):ChatGPT生成的内容,面临真实性、准确性以及版权方面的问题。这也给我们带来很多挑战——除了辨别真伪、保护版权,还要能够识别这到底是不是个人的创作,或者能区分出哪部分是AI生成,哪部分是人的创作。

陈雨强(第四范式首席科学家、联合创

始人):对于一篇专业性或学术性的文章,署名作者是一定要负责的,而恰恰ChatGPT是不能负责的。学术论文的署名作者,既要论文做出了“实质性贡献”,又要能承担相关的责任,而这两条ChatGPT都不满足。

何俊良(上海交通大学计算机系助理教授):如果问ChatGPT我们自己不熟悉的

专业领域的问题,它的回答会看起来非常专业,但实际上可能包含很多错误的知识和误导性观点,而我们难以分辨。

真实性问题会带来信赖危机,而一旦公众不敢信赖一款应用,它也便难于发展。所以,大语言模型目前最亟待解决的问题之一,就是怎么去判断AI的输出到底是对是错,学界也特别关心。

AI的普及给教育体系带来哪些挑战?

沈阳:有人做试验,让ChatGPT参加考试,其能达到硕士的入学水平,这预示着一般性脑力劳动在未来的价值会大幅降低。如果ChatGPT发育得比较充分的

话,我们未来的整个教育体系是要改变的。我们要培养下一代增加创造性知识的能力,提供AI所不能提供的价值。

唐明(中国科学院自动化研究所研究

员):我认为,ChatGPT能极大提高人类的工作效率,但所谓取代人类,则远远谈不上。AI再怎么发展,最终还是需要人做出判断,不可能完全信赖机器。据新华社