



科幻电影《机器人瓦力》中,瓦力开始具有自我意识,会好奇、会孤独、有梦想,最后还产生了爱情。

# 我们需要给人工智能上道德课

今年1月,在美国加州阿西洛马召开的“阿西洛马会议”上,包括844名人工智能和机器人领域的专家基于对人工智能和机器人迅猛发展的深切关注,联合签署了《阿西洛马人工智能原则》,呼吁全世界的人工智能领域的工作者遵守这些原则,共同保障人类未来的利益和安全。

在这总共23条关于人工智能和机器人的“军规”当中,关于必须让它们更有道德的条款格外引人注目。越来越多的科学家开始认为,无论是听话的还是不听话的人工智能,对于人类来说都有危险性,只有和人类具有一样价值观,人工智能才能真正成为人类的好朋友。

## 听话的机器人VS不听话的机器人 哪个更安全

人工智能通常被认为是冷血的。比如在2015年的一部科幻电影《机械姬》(Ex Machina)中,机器人“艾娃”(Ava)诱骗了一个年轻男子,让他帮助自己杀死了创造自己的纳森(Nathan)。“艾娃”的行为在人类看来是邪恶的,但换个角度来审视,或许也是某种正义:正如被杀死的纳森一开始就想到的那样:“有一天,人工智能机器看向我们时,就像现在的我们看着非洲平原上的化石残骸。它们眼中只会看到一种说着粗鲁语言,用着低级工具的直立行走着的猿类——一个注定要被淘汰的物种。”

出于这种担忧,很多研究机构已经在未雨绸缪,比如,谷歌的人工智能机构正在研究一个能够让人工智能软件停止运作的“红色大按钮”。这是一个关闭人工智能的开关,但它并不像我们的家用电器的开关键这么简单,因为人工智能本身有深度学习的能力,这就意味着如果它认为逃避人类的打断可以增加预期收益,它就会想办法阻止人类“关机”——而谷歌想要做到的是,探索“一种确保机器人不会试图防止(或寻求)来自环境或人类的打断的方法”。

但这其中似乎存在一个悖论,因为很多时候,机器人又必须学会对那



科幻电影《机械姬》中,机器人“艾娃”不择手段杀死了创造自己的人。

些无法完成或可能伤害自身或违反道德的命令说“不”。否则,机器人及人工智能技术会放大人为错误,或被一些不法分子利用。老问题又出现了:当聪明的机器人被赋予深度学习能力,甚至能够轻而易举地发现人类犯下的错误时,他们是不是会走上反抗人类的道路?

## 讲道德的机器人最安全

《阿西洛马人工智能原则》被视为是对人工智能和机器人的发展而产生的对人类难以估量的影响的长期观察、思考的一次集中性表达。

在这23条关于人工智能的“军规”当中,有一条规定特别引人注目:人工智能系统应该被设计和操作为与人类尊严、权力、自由和文化多样性的理想相一致。换句话说,人工智能具有正确的“三观”是非常有必要的。

北京大学心理与认知科学学院教授朱滢表示,实际上,人工智能已经足够聪明威胁到我们,就好像病毒没有意识,也没有智力,但它对人类十分危险是一个道理。它危险不是因为人工智能有意识,只是因为它具有超高的智力,同时又不具备恰当的伦理观念,或者说——没有道德造成的。

上海交通大学科学史和科学文化研究院副教授杜严勇表示,对人工智能产品进行伦理设计是解决其安全问题的基本进路之一。机器人伦理研究的主要目标,就是让机器人在与人类互动的过程中,具有一定的道德判断与行为能力,从而使机器人的所作所为符合人们预设的道德准则。

## 三种模式实现人工智能的道德抉择

美国学者瓦拉赫(Wendell Wallach)和艾伦(Colin Allen)认为,将来人工智能系统必定会独立于人类的监管,自主做出决定,他们把能够做出道德抉择的人工智能系统称之为人工道德行为体(artificial moral agents,简称AMAs)。他们为人工智能的道德抉择设计了三种实现模式:“自上而下的进路”、“自下而上的进路”以及“混合进路”。

自上而下的进路是指选择一套可以转化为算法的道德准则作为机器行为的指导原则;自下而上的进路类似于人类的道德发展模式,通过试错法培养道德判断能力。不过,这两种方式均有一定的局限性。比如,第一种进路把一套明确的规则赋予机器可能是不合理的,同一种原理在不同的情况下可能会导致不同的相互矛盾的决定。后一种进路要求机器能够自我发展进化,而人工智能系统的学习与进化能力的提高将会是一个比较缓慢的过程。混合进路在一定程度上,可追溯到亚里士多德的美德伦理学,它把两种进路统一了起来:一方面,美德本身可以清楚地表述出来;另一方面,它们的习得又是典型的自下而上的过程。

虽然要真正实现对人工智能的伦理设计还有许多具体的工作要做,但是从机器人伦理研究的勃兴及取得的初步成果来看,在人工智能产品中实现人工道德完全是可能的。人们也更容易相信,一个能够根据我们的预设做出合理的道德判断和行为的人工智能系统,对我们来说才是真正安全可靠的。

摘自《广州日报》

## ■看天下



## 大熊猫爱宝乐宝韩国庆生

大熊猫爱宝和乐宝在韩国最大主题公园爱宝乐园里迎来了在异国他乡的第二个生日。据了解,爱宝(雌性)出生于2013年7月13日,乐宝(雄性)出生于2012年7月28日。

庆生活动现场,工作人员为这对小寿星分别制作了特别的生日蛋糕,让它们尽情享用。为乐宝准备的是用水果、竹叶和冰块制成的“果冰蛋糕”,清凉解暑;为爱宝准备的则是用新鲜竹子做成的4层大蛋糕,上面还插着胡萝卜蜡烛。不仅如此,工作人员还为熊猫们送上了冰镇西瓜,一消初伏的暑气。



## 机器人“护工”照顾老人

近日,日本汽车制造商丰田分享了一则视频,曝光其研制出的新款机器人。这款机器人1.22米高,有着夹子般的手臂,轮式底座,可以在平坦的地面上移动。视频中,使用者可以用平板电脑控制机器人端茶倒水。丰田将这款机器人称为“支援机器人”,意为专门用来帮助残疾人实现简单生活自理工作的机器人。

据了解,在养老压力巨大的日本,2013年政府便指定24家企业、拨款23.9亿日元补贴开发护理机器人,用于帮助老年人在房间内移动、如厕及报告行踪。



## 找到“天使粒子”

国际权威学术期刊《科学》杂志21日报道了物理学领域的一项重大成果:以华人科学家为主体的科研团队找到了正反同体的“天使粒子”——马约拉纳费米子,从而结束了国际物理学界对这一神秘粒子长达80年的漫长追寻。

该成果由加利福尼亚大学洛杉矶分校何庆林、王康隆课题组和美国斯坦福大学教授张首晟课题组、上海科技大学寇煦丰课题组等多个研究团队共同完成,论文通讯作者为何庆林、寇煦丰、张首晟、王康隆。图为张首晟。