

随着人工智能和机器学习的发展,机器人越来越多地承担起更加重要的工作,如外科手术、自动驾驶、战场作战,甚至能决定人类的生死,那么——

# 我们能信赖机器人吗?

本报记者 常丽君 综合外电

按大众对机器人的理解,它们要么是近乎完美的忠诚管家,要么是精神变态的狡诈杀手,就像《星球大战》里的C3PO、《机械姬》里的艾娃、《禁闭行星》里的罗比或《太空漫游》里的哈尔。虽说这些描写只不过反映了人类自己的希望和恐惧,但至少这些恐惧已开始进入现实。

近年来,许多科技界名人表达了他们对人工智能横行世界的忧虑:拥有超级智能的机器人建立了一个新世界,人类不再是这个新世界的主角,他们被奴役、杀戮甚至灭绝。这些恐怖场景和几十年前科幻小说家们幻想的差别不大,却引起了人们的广泛关注,甚至包括斯蒂芬·霍金、比尔·盖茨、埃隆·马斯克等科技名人。

## 精准高效的外科手术机器人

目前,外科医生已可以利用机器人手臂来做复杂的手术。美国纽约大学机器人手术中心主管迈克尔·斯蒂夫曼已做过上千例机器人辅助手术。他在控制台操纵着机械臂,每个机械臂都通过一个约5毫米宽的微小切口伸入病人体内,只要旋转自己的手腕,捏紧手指,伸入病人体内的机械臂就会精准地执行同样动作。

他引导着两只手臂给一根线打结,操纵第三只手臂用针穿过病人肾脏,将切除肿瘤后留下的洞缝在一起,第四只手臂则拿着内窥镜将病人腹腔内的情况呈现在显示屏上。

斯蒂夫曼是位受过高等训练的专家,有高超的技能和判断,然而,他正在把宝贵的时间花在缝合上。这只是主手术的后续整理工作,如果机器人能接手这项单调机械的任务,外科医生就能腾出手来做更重要的事。

今天的手术机器人能力进一步增强,手术中不仅没有手颤,还能实施多种方案。但说到底,机器人只是由人类直接控制的高级工



四臂手术机器人

## 进入实用的“达·芬奇”系统

2013年,位于加利福尼亚州的“直觉手术”公司开始向世界各地机器人研究人员捐赠“达·芬奇”手术机器人系统,每套造价高达250万美元,它也是美国监管机构批准的软组织手术系统。目前,全世界有超过3600家医院安装了“达·芬奇”系统。但它的商业之路并非一帆风顺,曾因轻微事故而面临诉讼。尽管充满了争议,很多医院和病人已经接受了这种技术。

“达·芬奇”处在人类医生的完全控制之下,除非医生在控制台上抓住操纵杆,否则其手臂只是毫无生气的塑料和金属片。该公司先进系统研发部经理西蒙·迪迈奥说,目前他们打算保持这种现状。但机器人专家正在向未来努力,让医生在做手术时有越来越多的辅助工具或计算机引导。

迪迈奥指出,这一领域的研究就像早期的自动驾驶汽车,“第一步是识别道路标志、障碍物、汽车和行人,下一步是让汽车来帮助司机。比如,智能汽车会感知周围车辆的位置,在司机错误变道时发出提醒。同样,手术机器人可在医生的手术器具偏离常规路径时发出警示。



▲手术机器人达·芬奇。▲加州大学伯克利分校机器人专家肯·戈德伯格让达·芬奇机器人缝合手术切口,达·芬奇能通过“机器学习”软件自己学习。

人类和机器人正形成一种新的关系。在不久的将来,我们会开始对那些具有高度自主性、完全自动化的机器人系统委以重任,如驾驶汽车、执行手术,甚至在战争中选择何时使用致命性武器。这是人类第一次在面对复杂、变动和无序环境时,通过对机器人编程而不是直接控制来做出生死抉择。

毫无疑问,机器人并不完美,它们也会犯错,甚至可能造成人类伤亡。但可以确定的是,这一新兴科技将给人类带来前所未有的发展前景。届时,我们将面对的挑战涵盖技术、法规甚至哲学范畴,除了法规和伦理问题,新的机器人还迫使我们面对更深的道德难题,甚至改变对自我的认识。但最终,这个世界会因为机器人时代的到来而变得更加美好。

具。手术机器人公司泰坦医疗(Titan Medical)执行副总裁丹尼斯·福勒曾是一名从业32年的外科医生,他认为,如机器人能代替人类自动做一些决策,独立执行分配给它们的任务,医疗行业将能更好地为人类服务。“这种技术干预手段能增加可靠性,减少人为错误。”

给机器人“升职”并非遥不可及的事,所需的大部分技术很快会在研究和工业实验中开发出来,实验机器人用人体组织的橡胶模型练习缝合、清洗伤口、切除肿瘤等。在一些实验中,机器人能比得上人,有些甚至比人更加精确、高效。就在上个月,华盛顿一家医院展示了一个机器人系统缝合的猪小肠组织,让人执行同样手术进行比较发现,机器人缝得更均匀、更细密。

虽然这些系统还没准备好用于病人,但它们代表了未来手术的发展方向。同样的逻辑也适用于操作室和装配流水线,如果高自动化能提高工作效果,那就没什么能阻止它。

肥胖症治疗医生、伦敦帝国学院讲师休丹·阿什拉芬对机器人手术的结果进行了研究。他认为,在可预见的将来,手术机器人将会在医生指令下处理简单的任务。“我们的目标是提高术后效果。如果用机器人能挽救生命,降低风险,那用这种设备就是义不容辞的。”

展望未来,医学界最终会使用下一代机器人——有决策权的人工智能。这种机器人不仅能处理常规任务,还能接管整个手术。虽然现在看起来还不太可能,但技术创新会自然而然地把人们引领到那里。阿什拉芬说:“这是一步步实现的,虽然每一步都不是特别大。正如50年前的医生想象不出现在手术室的样子,从现在起再过50年,估计又是另一番景象。”

其实手术机器人已经能自己做决定,独立性比人们想的更强。比如,在视力矫正手术中,机器人系统切下病人一小片角膜,并通过一系列激光脉冲重塑其内层;膝关节置换手术中,用自主机器人切割骨骼比医生更精确;在头发移植手术中,智能机器人能识别并采集病人头部坚固的头发毛囊,然后在秃斑头皮部位打下精密小孔,帮医生省下许多单调、费时的重复劳动。

那些涉及到胸腔、腹腔和盆腔部位的手术要面对更复杂的挑战,因为每个病人的生理结构都有所不同,自主机器人必须能很好地识别各种又湿又软的内部器官和血管;而且在手术中,病人的各部分内脏都可能起变化,所以,机器人还要能不断地调整手术计划。

机器人还要有可靠地处理危机的能力。比如,在肿瘤切除手术中突遇大出血,它必须能及时、正确地处理。手术中会出现各种不可预期的复杂情况,机器人成像系统和计算机视觉首先要能识别,并通过喷出红色液体提示这种情况的严重性;下一步,其决策软件要能找出处理这种问题的最佳方案,然后指令系统迅速将这一方案付诸行动;最后由评估程序来评价结果,确定是否还需进一步行动。让手术机器人掌握感知、决策、行动和评价每一步骤,代表着对工程师的巨大挑战。

加州大学伯克利分校自动化科学与工程实验室主管肯·戈德伯格也在训练他的“达·芬奇”独立执行手术任务。目前,它的缝合技术已相当灵巧,它能一只手拉线通过模型伤口两边,另一只手通过拉针把线绷紧,无需人类引导而开始缝下一针。它通过位置传感与摄像技术计算每次进出针的最佳位置,计划并跟踪针轨迹。但任务还是很艰巨,据报告,它目前只能完成四针程序的50%,而且还不会给线打结。

戈德伯格说,现在他们使用机器学习算法采集视觉和运动学数据,把每一针分为定位、推针等多个步骤,让“达·芬奇”按顺序处理。通过这种方法,它有可能学会做任何手术。

理论上,同样的程序也能指导真正的手术。戈德伯格认为,在今后10年内就能实现简单的手术任务自动化。但即使机器人真的能在常规手术中表现更好,他仍希望机器人的行动是一种在人类医生“监督下的自主”。他说,让机器人做长时间精确且一致的工作,就像缝纫机对手工缝纫,只有机器与人合作才能成为超级医生。



## 被重新定义的军事机器人

捷克作家卡雷尔·恰佩克1920年发表了科幻剧作《罗素姆的万能机器人》,发明了“机器人”一词,其最初含义是合成人类,它们在工厂长时间辛苦地工作,生产低成本商品。然而,最终这些机器人杀死了人类。这也是科幻小说一直在描写的情节:机器人失去控制,变成不可阻挡的杀人机器。

现在,随着人工智能的发展和机器人技术的进步,加上伊拉克、阿富汗战争中无人机与地面机器人的广泛使用,人们在震惊之余开始担忧,科幻中的恐惧会变成现实。

世界上最强大的军队目前正在开发更智能的武器,这些武器将拥有不同程度的自动化能力和杀伤力,绝大多数通过人类远程控制。但也有有人认为,未来的人工智能武器最终会完全自动化地操作,由集成微芯片和软件来决定人类的生死,这将成为战争史上的一道分水岭。

于是,“机器人杀手”引发了激烈争论:一方认为,机器人可能开启一场世界大战摧毁人类文明;另一方则认为,这些武器是一种新型精确制导武器,只会减少而不是增加伤亡。人工智能领域的一些主要研究人员呼吁,禁止“自主进攻型武器超越人类控制”。

去年,三位学术界名人——美国加州大学伯克利分校的斯图尔特·罗塞尔、麻省理工学院的马科斯·泰格马克和澳大利亚新南威尔士



▲美国魔爪SWORDS机器人 ▲进攻型无人机和射击机器人

## 悄然兴起的智能化武器装备

自主武器几十年前就已出现,但数量较少,而且基本上都是用于防御性目的。最近,军事供应商开发出了被认为是进攻型的自主武器。以色列航空工业公司的Harpy和Harop无人机,能朝着敌方防空系统发出的无线电波飞去,摧毁它们。公司称,这种无人机已广泛向全世界出售。韩国国防承包商DoDAAM Systems公司也开发了一种警卫机器人“超级eGis II”,装备有机枪,能利用计算机视觉自动探测射击3公里内的目标。据韩国军方报告,已经在临朝鲜边界的非武装地区对这些装备了机器人的武器进行测试。DoDAAM Systems公司称,他们已经售出超过30套这种系统,有些买家在中东。目前,这种自动系统的数量大大超过机器人武器。

有些分析师认为,未来几年,武器的自主程度将变得越来越高。位于华盛顿特区的一个无党派研究机构“新美国”的机器人战争专家彼得·辛格说:“战争将会变得完全不同,自动化将发挥重要作用,速度成为关键。”他预测,未来战争的战场就像无人机之间的混战,或机器人战舰和敌方潜艇之间的遭遇战,能提供到那间优势的武器将决定战争的胜负。“可能是高密度的直接对抗,对人类来说根本来不及介入,因为只在几秒钟内一切就已发生了。”

美国军方在一份无人系统路线图内,对这种新型武器列出了一些详细计划,但它在武器化方面的意图仍含糊不清。国防部长罗伯特·沃克去年3月曾在一次论坛中强调,要在AI和机器人方面投资,并称在战场上出现越来越多的自动化系统是不可避免的。

问及关于自主武器时,沃克坚持美国军方“不会把做致命决定

大学的托比·沃什,在一次人工智能(AI)大会上组织发起了联名请愿。他们在公开信中指出,这些武器将导致“全球AI军备竞赛”,用于“暗杀、破坏国家稳定、镇压群众和选择性地消灭某个少数民族群体。”目前,这封信已有超过两万人签名,包括著名物理学家斯蒂芬·霍金、特斯拉CEO埃隆·马斯克等人。马斯克还捐了1000万美元给一家位于波士顿的研究所。该研究所的任务是“保卫生命”,抵抗可能出现的恶意人工智能。这一事件成为世界各大媒体的新闻,甚至在12月日内瓦联合国裁军谈判会议上,约有100多个国家参与讨论这一问题。

这场争论还扩展到网络上。人们对未来进行了各种预测和展望,有人认为,某种情况下还可能出现“出售杀伤性微型机器人的低成本大规模黑市,购买者可以给它设定某些标准,不加分辨地杀死成千上万符合标准的人”。

他们三人还指出:“自主性武器是一种潜在的大规模毁灭性武器。虽然一些国家或许不会选择将它们用于这种目的,但这对有些国家和恐怖分子却极具吸引力。”

造出在智能、自主和自动化方面不断升级的杀人机器,不管这种军备竞赛能否更好地服务于人类利益,也不管目前存在着多大争议,实际上新一轮的AI军备竞赛已经开始。

的权利授权给一台机器。”但他也补充说,如果“竞争对手”更愿意把这样的权利交给机器的话,我们也不得不决定,怎样才能最好地竞争。

在开发陆海空无人战斗系统方面,俄罗斯也遵循同样策略,但至少现在还依靠人类操作。俄罗斯的M平台是一种小型遥控机器人,装备有卡拉什尼科夫步枪和榴弹发射器,类似美国“魔爪SWORDS”系统(配备M16及其它武器的地面机器人)。俄罗斯还造了一个更大的无人战车Uran-9,配备30毫米机关炮和反坦克导弹,去年还演了一个人形战士机器人。

联合国一直坚持讨论致命自主机器人已近5年,但其成员国未能起草一份协议。2013年,联合国人权问题特别报告员克里斯·海恩斯写了一份有影响的报告指出,世界各国有一个难得的机会,在自主武器大规模发展之前讨论它们的风险。

今年12月,联合国“特定常规武器公约”将召开5年审查会议,致命性自主机器人的话题将提上日程,但要在会议上通过禁令是不可能的。这一决策需要所有参会国一致同意,在如何处理未来可能出现的泛自主武器问题上,各国仍存在根本分歧。

最终,“杀手机器人”的辩论似乎更关乎人类自身而不是机器人,自主武器和其他任何技术一样,至少在第一次使用时,应该更加谨慎,否则会是混乱和灾难性的。像“自主作战机器人是个好主意吗?”这样的提问可能不太合适,更好的应该是“我们确信自己能充分信任机器人并和它们生活在一起吗?”

## 兼顾逻辑与道德的无人驾驶车

设想一下,将来的某个夜晚,一个醉醺醺的行人突然在一辆无人驾驶汽车前面摔倒,结果当场被撞死。如果车里有人,这会被认为是一次事故,因为行人显然有错,理性的司机也难以及时躲避。但到了本世纪20年代,随着无人驾驶汽车日益普及,车祸发生的概率会减少90%,对司机过失的“理性人”的法律标准也会消失,对应的标准是给“理性机器人”。

死者家属会把汽车制造商告上法庭,指称汽车虽然来不及刹车,但它应该躲避周围行人,越过双黄线,撞上旁边车道上那辆空空的无人驾驶车。而根据该汽车的传感器数据分析,再现撞人场景也确实如此。原告律师会质问汽车头部软件设计师:“为什么汽车不会转弯?”

但法庭绝不会质问一个司机,因为那将撞车前他不采取某个紧急措施,因为这是毫无意义的——司机惊慌失措无法思考,只能凭本能做事。但如果司机是个机器人,问“为什么”就变得合理了。

在人类的道德标准、有瑕疵的典例、各种各样工程师想不到的假设情况中,最重要的假设是:一个有着良好判断能力的人知道何时应不顾文字上的法律条文,真正维护法律精神的至高无上。现在工程师必须做的是,把一些基本的判断因素教给自动驾驶汽车和其他机器人。

目前,在英国部分地区、德国、日本、美国4个州和哥伦比亚特区,法律已明确允许对完全自动化车辆进行测试,但车里还要有一个测试司机。谷歌、日产、福特及其他公司也表示,他们预计在今后5到10年内实现真正的无人驾驶操作。

自动驾驶获得环境信息是通过一系列传感器,如视频摄像头、超声波传感器、雷达和激光雷达(激光测距)。在加利福尼亚州,申请自动驾驶牌照要向机动车管理局提供碰撞前30秒的所有传感

器数据,工程师可以凭着这些数据精确重现碰撞事故场景。利用汽车的传感器记录,就能推知它决策背后的逻辑。发生事故后,监管机构和律师能根据这些记录,坚称自动驾驶车辆有着超人的安全标准,通过了严格的审查。制造商和软件开发商也将以目前人类司机无法想象的方式,为无人驾驶车的行为辩护。

所有驾驶都会涉及到风险问题,但这些风险如何在司机、行人、骑自行车者之间分配,甚至风险的性质都有着道德的成分。无论对工程师还是对一般公众来说,最重要的是自动驾驶汽车的决策系统,它决定了汽车行为中所含的道德成分。

对于道德上模棱两可的情况,在遵循法律的同时尽量减少损失。这种策略很有吸引力,因为它允许开发人员为肇事车“不作为”的行为辩护,也向立法者传递了定义道德行为的责任。但不幸的是,这方面的法律尚未完善。

在大多数国家,法律依赖于人类的常识,对自动驾驶汽车编程也应遵守法律规定:不得越过双黄线,即使有撞上一个醉汉的风险,即使另一边车道上只有空的无人驾驶车——法律很难为紧急情况制定例外,汽车开发商无从确定何时越过双黄线是安全的。

在道路车辆自动化伦理学上,这并非不可解的难题。可以借鉴其他领域的大量案例,来安全合理地处理类似的风险和利益。比如,给病人捐献器官,依据它会带来更有质量的生命还是残疾的生命,还增加了一些对社会更有用的职业。自动驾驶车辆面临的更大挑战是,它们必须在信息不完全、程序员常常考虑不到的情况下,利用编码在软件中死板的道德条款,迅速做出决定。

幸运的是,公众并没有过分地期望超人智慧,考虑到伦理道德的复杂性,只要对自动驾驶行为有个合理的判断就可以了。一种解决方案不一定要完美无瑕,但应该是多方考虑、合情合理的。