2024年8月2日 星期五
编辑/王文洁 投稿邮箱：wlsk@kepu.gov.cn 科普时报

对机器人该有怎样的伦理约束

□ 尹传红

第一，机器人不得伤害人类，也不得坐视人类受到伤害而袖手旁观；

第二，在满足第一定律的情况下，机器人必须服从人类的命令；

第三，在满足第一定律和第二定律的情况下，机器人必须保护自己。

——阿西莫夫“机器人三定律”

第二十六届上海国际电影节“一带一路”电影周期间，套开了一个以“AIGC时代的科幻创意与技术变革”为主题的第六届上海科幻影视产业论坛，探讨的是人工智能生成内容（AIGC）如何改变科幻影视产业，其中又设立了“机器人伦理道德与公共安全”的分论坛。

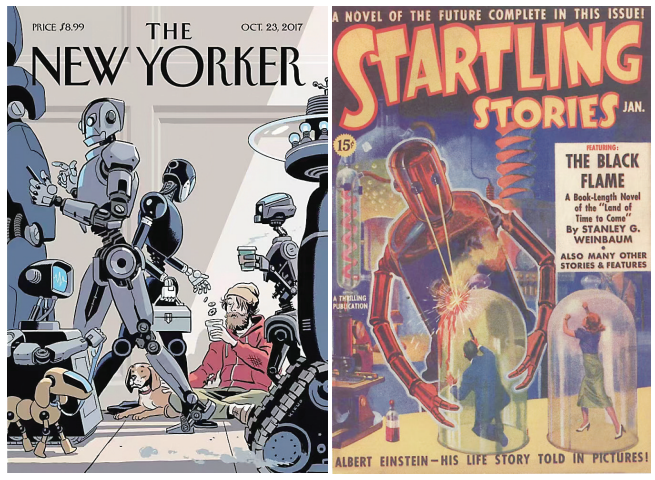
这个分论坛实际上是五个人对谈。身为科幻作家的主持人竹君开场就提出了一个关于机器人的核心问题：长期以来被研究机器人和人工智能伦理的人奉为经典的阿西莫夫“机器人三定律”，是否仍适用于解决当今的机器人伦理与道德问题？

首先回答这个问题的豪微研究院院长孔华威指出，现在的人工智能和机器人技术已经超越了阿西莫夫科幻小说中的想象，需要新的伦理框架来约束和指导机器人行为。

我在回答问题前作了一些补充介绍：阿西莫夫“机器人三定律”的完整表述，最早出现在阿西莫夫1941年创作的科幻小说《环舞》（又译《转圈圈》）中。这三条定律构成了支配机器人行为的一套“道德标准”，从而给“机器人社会”赋予了新的伦理。阿西莫夫通过这三大定律的相互作用，构思出一系列情节紧张、妙趣横生的短篇科幻小说《我，机器人》。好莱坞曾根据阿西莫夫机器人科幻小说改编拍摄过两部著名的科幻电影《机器管家》和《机械公敌》。人们也尽可想象，现实科技中它们会作为准则被编成程序，输入到计算机化的机器人大脑中去，成为研制和使用机器人必须遵循的基本法则。

“机器人三定律”的制定，反映了人类对机器人的恐惧心理。出于保护人类自身安全的考虑，对这种“人造机器”，必须要有明确的程序和技术手段来限制其行为。不过，在今天复杂的技术环境中，“机器人三定律”显得过于理想化了。确保机器人行为的安全和可控，需要更复杂的伦理和法律框架来规范，自然也要有更先进的技术介入。

AI+海洋科创中心常务副主任赵辉提到，“机器人三定律”在实际应用中常常被忽视，例如在战场上，无人机执行的是人的指令，而不是遵循“机器人三定律”。科幻作家江波谈及，“机器人三定律”是一种美好的愿景，但在实际的技术实现中，存在许多难以克服的挑战。目前的技术和伦理讨论，应该更加关注如何在现实中实施有效的控制和监督机制，以规避不可控的风险。



左图为2017年10月号《纽约客》杂志封面图。它描绘的是人类坐地行乞的场景，机器人则扮演了施予者的角色，折射出人们对未来的焦虑。

右图为美国1953年出版的一本科幻杂志的封面图，展现的则是人与机器怪物争斗的场面。

（作者供图）

讨论还涉及到机器人未来可能具备的自主意识。确实，现代科技术语“涌现”的提出、复杂性科学的发展，大大扩展了人类的认知边界，也启迪我们思考类似这样的问题：机器人或人工智能体是否能够拥有自由意志或自我意识？阿西莫夫1976年发表的机器人科幻名篇《活了两百岁的人》，对此做了非常精彩的想象和描述。

美国计算机专家和科幻作家弗诺·文奇、发明家和未来学家雷·库兹韦尔等学者则早在20多年前就关注到“奇点”话题，并把它当作一个“隐喻”：由超人类智能驱动的进步异常迅猛，并将达到一种临界聚集状态。当智能机器的能力跨越这一临界点后再加速喷发，有可能就会突破任何试图对它进行的管控。

也有这样一种可能：未来的机器人通过大数据与算法自行学习和进化，并给人类社会带来不可预见的伦理挑战。例如，机器人可能会在无人监管的情况下，进行自我优化和升级，从而发展出与原始设计意图完全不同的模式。这种能力很可能会超出人类的控制范围。未来技术的核心，或许是通过计算与大数据来管理和规范机器人行为。

要警惕陷入这样一种“技术困境”：当我们不期然跨越了某种技术的门槛、越过了社会限定的红线，才认识到我们不希望的后果终究出现的时候，技术已经深度融入我们的社会经济生活，我们除了无奈接受，对它的控制也变得尤其困难了。

隐形卫士

□ 十七

鸥城简直是一座绿岛，远远看去，各色房子像是浮在碧浪里。

粟洁穿过院内的翠绿长廊，脚步轻快，出院门左拐50米，来到鸥城中路蜀山公交车站，364路公交车还没到。蜀山公交站附近有十七八个居民小区，共有27个线路的公交车停靠，人来车往，十分繁忙。

根据数据中心信息分析的结果，粟洁今日被分派到364路公交车上，她将在终点与起点之间不停换乘，目的是要找到一个人。这无异于大海捞针。

粟洁选了公交车上距离车门不远的座位，刚坐下，就有一位小伙子过来搭讪。粟洁身材高挑，人海中总是最吸睛的那一个。

粟洁明白小伙子的意图，主动出示了自己的微信二维码，让他加了微信。然后，粟洁继续专注盯着上车、下车、站台上的每一个人，希望尽快找到目标人物，消除隐患。她要找的那个人很狡猾，在碳基生命中，他的智商确实很高，安防系统曾多次跟丢目标。粟洁对他了解也不多，因为信息是分层管理的，层级不一，信息获取量就不一样。

公交车平稳驶过市区，终点站是栖凤岭工业区。工业区有许多小微企业，他们没有通勤车，364路车回程时往往爆满，地铁就更不用说了。

今日，整个鸥城无数双眼睛得到的信息实时汇总，经数据中心分析后，得出的结果实时分发出去，可那人却像是人间蒸发了。

下午3时许，粟洁在栖凤岭下车，换乘另一辆364路公交返回城里。下班的人群蜂拥而上，粟洁被挤靠在车门上动弹不得。公交车启动后，她才有点空隙站稳脚跟，然后通过车载摄像头，对车上的人进行过滤。

突然，她的心率加快，第六感告诉她，那人就在车上。但她必须精准确认，再对他定位，并判断潜在的危险指数。车厢里拥挤乘客让她的行动稍微费了些劲。

那人离她约有一米远，在公交车中门靠后的位置，携有一包。粟洁费尽周折也看不清包里的东西。那人反侦察能力很强，他的包带有离子屏蔽功能。

粟洁不得不连接上后台主机，这样一来，她对身边的态势感知就会大打折扣。好在连接主机用时不长就有了结果——他包里有一部手机、一包烟、一只无烟打火机和两本书，还有一个长方形的片状物体，主机识别也很模糊，被定性为可疑危险品。

公交车进城后，速度慢下来，走走停停，车上的乘客越来越少。这时，粟洁得以与那人面对面。经过再次确认，是目标人物无疑。粟洁掏出手机，背对那人玩起了游戏，并撕开一粒口香糖，咀嚼着。

离蜀山还有一站时，那人拎着包下了车，向公交车的相同方向走去。

粟洁没有下车，通过布放出去的蜱虫追踪器跟踪那人。临近蜀山车站时，那人在街边站了一会儿，然后快步穿过马路，跑到街对面蜀山车站。粟洁立即搜索蜀山车站周边车流人流信息，发现有两路公交车即将靠站，公交车后不远，一辆校车正徐徐驶来。

蜀山站挤满了人，粟洁调取人脸数据比对，发现站台上大多是接孩子的家长。那人站在人群里，表情木然。

信息飞速上传后，总部发来的是特级警讯。

不容迟疑，粟洁迅速穿过车流，向那人冲去。那人见状，从包里拿出一个极薄的盒子，抽出一把尺子长短的陶瓷刀，对冲到面前的粟洁一阵猛刺乱砍。人们像炸了窝的马蜂，四散奔逃。这时，满载孩子的校车驶进蜀山站。

那人一见校车，立即凶相毕露，撇下粟洁冲向校车，孩子们被吓得哇哇大叫，乱成一团。粟洁抓住那人的胳膊，将他驱离校车。那人回转身体，乱刀在粟洁脸上、身上又砍又刺，发出刺耳的挫骨声。粟洁没有退缩，越战越勇。很快，那人被粟洁勒住了脖子，不一会儿就乖乖地束手就擒。

逃散开的人们从惊魂中醒来，赶紧跑回来帮助粟洁。粟洁的衣服多处被划烂，脸上、手上、前胸后背等多处被陶瓷刀划破，但没有血液流出，伤口下露出银灰色的钢铁身躯。

孩子们依偎在爸妈怀里向粟洁致敬。粟洁笑了，只是，脸部被划开的硅胶肌肉模糊了她原本迷人的笑容。

（作者系四川省资阳市作家协会会员）