新闻热线：010—58884062
E-mail：erin74mail@163.com
■责编 翟冬冬
科技日报 2018年11月5日 星期一

直面隐忧 业界大咖谈人工智能作恶

别“炼”出造反的AI

本报记者 **刘 垠**

一场抢劫案后，格雷的妻子丧生，自己也全身瘫痪。他接受了一个天才科学家的“升级”改造治疗——在他身体里植入了人工智能程序STEM，获得了超强的能力，从一个“残废”直接升级成为职业杀手。随着STEM的进化升级，步步紧逼格雷交出身体使用权和大脑意识控制权……

本年度关于人工智能和人类未来的最佳影片，不少人认为非《升级》莫属。而人工智能和人类抗衡的探讨，是科幻电影中的永恒话题，从《银翼杀手》到《机械姬》，再到今年的低成本电影《升级》，都映射出未来人工智能对人类的威胁。

黑产超正规行业 恶意源于人类基因

AI造反，是科幻电影里太常见的桥段。问题在于，现实当中真正的AI好像也在一步步向我们走来。不少人抱有忧虑和不安，人工智能会“作恶”吗？

倾向于AI威胁论的人并不在少数。马斯克曾在推特上表示：“我们要非常小心人工智能，它可能比核武器更危险。”史蒂芬·霍金也说：“人工智能可能是一个‘真正的危险’。机器人可能会找到改进自己的办法，而这些改进并不总是会造福人类。”

“任何技术都是一把双刃剑，都有可能用于作恶，为什么人工智能作恶会引起这么大的反响？”在近日召开的2018中国计算机大会的分论坛上，哈尔滨工业大学长聘教授邬向前抛出了问题，人工智能研究的底线到底在哪里？

早在1942年，阿西莫夫就提出了机器人三定律。但问题在于，这些科幻书中美好的定律，执行时会遇到很大的问题。

“一台计算机里跑什么样的程序，取决于这个程序是谁写的。”360集团技术总裁、首席安全官谭晓生说，机器人的定律可靠与否，首先是由人定义的，然后由机器去存储、执行。

值得注意的是，“不作恶”已成科技行业的一个技术原则。那么，机器人作恶，恶意到底从何而来？

如今人工智能发展的如火如荼，最早拥抱AI的却是黑产群体，包括用AI的方法来突破验证码，去黑一些账户。谭晓生笑言：“2016年中国黑产的收入已超过一千亿，整个黑产比我们挣的钱还要多，它怎么会没有动机呢？”

“AI作恶的实质，是人类在作恶。”北京大学法学院教授张平认为，AI不过是一个工具，如果有人拿着AI去作恶，那就应该制裁AI背后的人，比如AI的研发人员、控制者、拥有者或是使用者。当AI在出现损害人类、损害公共利益和市场规则的“恶”表现时，法律就要出来规制了。

目前，无人驾驶和机器人手术时引发的事故，以及大数据分析时的泛滥和失控时有耳闻。那么，人工智能会进化到人类不可控吗？届时AI作恶，人类还能招架的住吗？

任务驱动型AI 还犯不了“反人类罪”

值得关注的是，霍金在其最后的著作中向人类发出警告，“人工智能的短期影响取决于谁来控制它，长期影响则取决于它能否被控制。”言下之意，人工智能真正的风险不是恶意，而是能力。

“人工智能未来的发展会威胁到人类的生存，这不是杞人忧天，确实会有很大的风险，虽说不是一定会发生，但是有很大的概率会发生。”在谭晓生看来，人类不会被灭亡，不管人工智能如何进化，总会有漏洞，黑客们恰恰会在极端的情况下找到一种方法把这个系统完全摧毁。

对此，上海交通大学电子系特别研究员倪冰冰持乐观态度。“我们目前大部分的AI技术是任务驱动型，AI的功能输出、输入都是研究者、工程师事先规定好的。”倪冰冰解释说，绝大多数的AI技术远远不具备反人类的能力，至少目前不用担心。

张平表示，当AI发展到强人工智能阶段时，机器自动化的能力提高了，它能够自我学习、自我升级，会拥有很强大的功能。比如人的大脑和计算机无法比拟时，这样的强人工智能就会对我们构成威胁。

“人类给AI注入什么样的智慧和价值观至关重要，但若AI达到了人类无法控制的顶级作恶——‘反人类罪’，就要按照现行人类法律进行处理。”张平说，除了法律之外，还需有立即“处死”这类AI的机制，及时制止其对人类造成的更大伤害。“这要求在AI研发中必须考虑‘一键瘫痪’的技术处理，如果这样的技术预设做不到，这类AI就该停止投资与研发，像人类对待毒品般全球诛之。”

作恶案底渐增 预防机制要跟上

事实上，人们的担忧并非空穴来风。人工智能作恶的事件早在前两年就初见端倪，比如职场偏见、政治操纵、种族歧视等。此前，德国也曾发生人工智能机器人把管理人员杀死在流水线的事件。

可以预见，AI作恶的案例会日渐增多，人类又该如何应对？

“如果我们把AI当作工具、产品，从法律上来说应该有一种预防的功能。科学家要从道德的约束、技术标准的角度来进行价值观的干预。”张平强调，研发人员不能给AI灌输错误的价值观。毕竟，对于技术的发展，从来都是先发展再有法律约束。

在倪冰冰看来，目前不管是AI算法还是技术，都是人类在进行操控，我们总归有一些很强的控制手段，控制AI在最高层次上不会对人产生一些负面影响。“如果没有这样一个操控或者后门的话，那意味着不是AI在作恶，而是发明这个AI工具的人在作恶。”

凡是技术，就会有两面性。为什么我们会觉得人工智能的作恶让人更加恐惧？与会专家直言，是因为AI的不可控性，在黑箱的情况下，人对不可控东西的恐惧感更加强烈。

目前最火的领域——“深度学习”就是如此，行业者将其戏谑地称为“当代炼金术”，输入各类数据训练AI，“炼”出一堆我们也不知道为啥会成这样的玩意儿。人类能信任自己都无法理解的决策对象吗？

显然，技术开发的边界有必要明晰，比尔·盖茨也表示担忧。他认为，现阶段人类除了要进一步发展AI技术，同时也应该开始处理AI造成的风险。然而，“这些人中的大多数都没有研究AI风险，只是在不断加速AI发展。”

业界专家呼吁，我们必须清楚地知道人工智能会做出什么样的决策，对人工智能的应用范围和应用结果的预期，一定要有约束。

AI会不会进化，未来可能会形成一个AI社会吗？“AI也许会为了争取资源来消灭人类，这完全有可能，所以我们还是要重视AI作恶的程度和风险。”现场一位嘉宾建议，我们现在要根据人工智能的不同阶段，比如弱智能、强智能和超智能，明确哪些人工智能应该研究，哪些应该谨慎研究，而哪些又是绝对不能研究的。

如何防范AI在极速前进的道路上跑偏？“要从技术、法律、道德、自律等多方面预防。”张平说，AI研发首先考虑道德约束，在人类不可预见其后果的情况下，研发应当慎重。同时，还需从法律上进行规制，比如联合建立国际秩序，就像原子弹一样，不能任其无限制地发展。

好机友

清华AI画虾师 想当现代齐白石

图1为“道子”所画

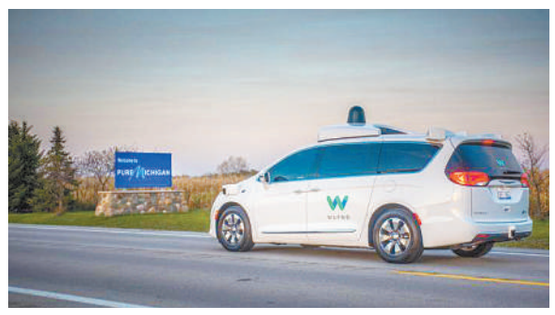
据量子位报道，在近日播出的某综艺节目中，清华大学智能中国画系统“道子”火了。“道子”的过人之处在于，它能够学习不同画的绘画风格，将眼前看到的景物，绘制成一幅具有特定风格的图画。

作为画坛“新人”，“道子”自然也向齐白石虚心“拜师”学习画虾，节目现场“道子”与两位人类专业画师同台竞技，让观众找出哪幅画为AI所作。最终，3位嘉宾和现场100名观众在两轮比赛中，还是没能成功将“道子”找出。

清华大学博士后高峰介绍，比赛现场，“道子”借助摄像头拍下鱼缸中大虾的姿态图，随后利用风格迁移法，将齐白石的画虾风格转移到这张图像，最后由他为画面整体布局，完成虾图。简单来说，高峰给“道子”的神经网络模型喂食了大量国画大家的作品，让它在齐老先生的虾海，徐悲鸿的马群、黄宾虹的山水中归纳总结，直到领悟出如何提取各种画作特点。

情报所

加州批准！ 谷歌获完全无人驾驶许可证

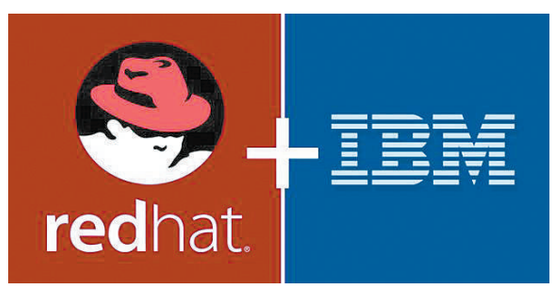


据路透社报道，美国加州机动车管理局近日表示，谷歌母公司Alphabet的自动驾驶公司Waymo成为第一家获得加利福尼亚州完全无人驾驶许可证的公司。此前有约60家公司包括苹果等公司获得自动驾驶汽车测试许可，但都有人类司机坐在方向盘后面。

加利福尼亚州官方表示，Waymo可以在圣克拉拉县使用大约30辆没有驾驶员的测试车辆。作为获得批准的一部分，Waymo必须持续监控测试车辆的状态，并与乘客提供双向沟通，且拥有至少500万美元的保险，并通知当地社区测试情况。

Waymo的许可证包括在城市街道，乡村道路和高速公路上进行日夜测试，其限速最高达到每小时65英里。该公司表示，测试车辆完全可以应对雨雾天气，并可以在这些条件下进行测试。

开源史上最大收购案 IBM340亿美元收购红帽



据路透社报道，近日IBM和红帽（Red Hat）共同宣布，两家公司已达成最终协议，IBM将收购红帽所有已发行的普通股，以每股190.00美元现金，总价值约340亿美元的价格正式收购后者。收购完成后红帽将被并入IBM的混合云部门。此次交易是IBM迄今最大一次并购。

IBM公司董事长，总裁兼首席执行官罗睿兰说：“收购红帽是一个改变游戏规则的方式。它改变了有关云市场的一切。IBM将成为全球排名第一的混合云提供商，为企业提供唯一的开放云解决方案。”

IBM成立于1911年。多年来该公司营收一直下滑，因此不得不从电脑制造业务转型为新技术产品及服务，最近开发的项目包括以其开发的超级计算机沃森命名的人工智能产品。红帽对于程序员来说是一个家喻户晓的名字，很多普通人也耳熟能详。不可否认的是，尤其是在云计算和Linux生态系统方面，红帽是一家重要的公司，拥有众多的业务。

外媒评价，这一举措对IBM来说意义重大。IBM和红帽“联姻”，或能挑战亚马逊、微软地位。

（本版图片来源于网络）

智慧交通，实现城市效能的最大化释放

第二看台

实习记者 **钱 力**

在不久的将来，你只需要手机点一点，就能召唤一辆巴士？乘客信息、路线规划、流量监控，都可以在系统内进行精准地响应？这一天，或许已经不远。

目前广州公交集团的大数据系统中每天能产生超过500亿条数据信息，涵盖乘客信息、各种车船状态信息，司机的行为信息，精确到急刹车、打哈欠等状态……人、车、生活，正随着科技的发展变得越来越紧密。

在近日召开的首届中国智慧交通大会上，交通运输部总工程师周伟表示：“新一轮科技革命和产业变革蕴育兴起，云计算、大数据、物联网、人工智能等快速发展，引发了以绿色、智能为特征的群体性技术变革。”

大数据引领交通运输业升级换代

去年年初，交通运输部发布了《推进智慧交通发展行动计划（2017—2020年）》（以下简称《行动计划》），明确提出围绕提升城际交通出行智能化水平、加快城市交通出行智能化发展等方面，推动企业为主体的智慧交通出行信息服务体系建设，促进“互联网+”便捷交通发展。

周伟透露，交通业近年正不断推进智能化技术创新。目前，首批三家自动驾驶封闭测试场地获得认定，国家智能网联汽车上海试点示范封闭测试区已建成200个智能驾驶测试场景。无人机在公路寻检、突发事件的现场监测已经在部分地区规模应用，无人机物流配送正在积极试点。国内首条全自动运行的地铁线也于去年年底在北京燕房线投入运营。

从数字化到智能化、智慧化，再到智慧网联，传统的交通运输业不断升级换代。“如果说信息化是改造了以往传统公共交通模式，那么现在的大数据则是引领了整个公共交通行业，是往精益化发展的新阶段。”深圳市地铁集团有限公司副总经理简炼说。

这一趋势从广州市公共交通集团的尝试中可见一斑。其副总经理张海燕讲到，“500亿条数据信息让我们能够更加优化资源、实现更符合市民出行需求的线网优化配置，同时对设备进行全寿命周期的跟踪管理，以保障司机和车辆的最佳状态。与此同时，通过分析用户的出行习惯，也能够为他们提供更加丰富的有关候车、实时路线轨迹等信息，乃至为用户提供定制化服务。”

智慧化的交通不仅仅能够给用户提供更加个性化、精准化的服务，更能够助力城市公共交通发展提升运行效率。腾讯今年复盘了深圳改革开放40周年灯光秀的热力图，发现8万人集中返回，交通疏导一定是个问题。而利用LBS技术和腾讯生态体系下产生的大数据连接交通管理部门，有效验证了城市现有公交线路规划的合理性，进行新公交线路的设计，实现对交通、人流的精准管理。

亟待实现双向开放破除数据“孤岛”

2017年腾讯车联推出“AI in Car”智能解决方案，共享其内容生态和服务生态。此番推出的系列智能交通产品，“智能出行助手”通过实时公交/地铁播报、线路规划，为用户提供全方位的出行服务，提升出行效率；“定制巴士”则根据用户需求以及客流情况定制巴士路线，多人成团，精准规划城市交通需求，提高服务效率，创造更多收益。

“智慧交通是智慧城市建设的核心，也是痛点之一。目前数据的碎片化所形成的分割，以及企业之间的壁垒等，容易造成一种‘孤岛’现象，很多数据不能互联互通和开放共享。”深圳巴士集团股份有限公司总经理王慧农讲到，有了共享的数据，才能使城市交通规划设计和服务模式更加智慧化、人性化。

目前，我国开放共享数据资源体系基本建立，通过政企合作方式建立了综合交通出行的大数据开放云平台，目前接入全国25家省市交通运输主管部门，43家包括腾讯、百度、高德、滴滴等在内的企业和机构，开放数据159项。信息服务接口140余项，推动了交通运输，公共信息资源的开放共享。

《行动计划》要求，各部门要加强交通运输大数据应用中心建设，建立交通运输政务信息资源共享机制，完善部省两级数据资源目录体系，推动综合交通运输基础信息交换共享。

“‘互联网+公共交通’的数字化升级并不仅仅在末端信息化。第一部分是场站、线网等交通基础设施，二是载运工具本身，三是我们的用户终端。需要做好三者的开放、协同与融合。”交通运输部公路科学研究院研究员孟春雷说。