科学向未来

光明图片/视觉中国

构建新一代人工智能准则

□ 曾 毅

著名作家阿西莫夫1940年在科幻小说中提出了“机器人三原则”:机器人不得伤害人类,或看到人类受到伤害而袖手旁观;机器人必须服从人类的命令,除非这条命令与第一条相矛盾;机器人必须保护自己,除非这种保护与以上两条相矛盾。今天,人工智能逐渐从科幻走向现实,人们对人工智能可能产生的危害也愈加警惕,希望能够为其制定准则以确保人工智能科技和产业向对社会有益的方向顺利发展。

1 以和谐互补的方式链接不同的人工智能准则提案

人工智能既是经过数十年锤炼发展出来的科学领域,又是改变未来的重要颠覆性技术。人工智能的研究与发展不仅关系到国家科技、经济发展、社会稳定,更关系到国家在科技、产业领域的国际影响与国际局势。

人工智能在带来机遇的同时也带来潜在的风险与隐患。例如应用最为广泛的深度神经网络模型在对输入引入极小的噪声(如改变图像输入中某个关键的一个像素值)情况下,就有可能使网络的识别和预测结果产生颠覆性的错误(如将青蛙识别为卡车、乌龟识别为枪支)。若没有充分的风险评估,新兴技术在对社会发展带来机遇的同时,很有可能引入不可预计的安全隐患与风险。

现在对于人工智能的发展而言,首要的问题是选择正确的道路。创新、价值、伦理是一个铁三角,创新性技术对社会带来潜在价值的同时,可能存在难以预期的风险,并对社会伦理提出重大挑战。因此,发展人工智能技术普惠经济与社会的同时,关注人工智能的社会属性,从社会风险、伦理准则与治理角度确保人工智能科学、技术、产业的健康、良性发展至关重要。

为了确保发展有益的人工智能,世界上各国政府、非政府组织、科学团体、科研机构、非营利性组织、企业都提出了人工智能发展准则。包括英国政府、国际电子电器工程学会、国际劳工组织。至今,于公开渠道可见的人工智能准则提案已接近40个,涉及以人为本、合作、共享、公平、透明、隐私、安全、信任、权利、偏见、教育、通用人工智能等主题。例如,美国生命未来研究所倡导的阿西洛玛人工智能准则(Asilomar AI Principles)、英国上议院提出的人工智能准则(AI Code)等,都希望通过在人工智能伦理与准则制定方面的领先来引领人工智能的发展。

事实上目前任何一个国家、机构、组织提出的人工智能准则都仅覆盖了少部分主题(更为具体的层面,人工智能准则提案涉及超过50个主要议题),虽然不少准则提案都有自身的特色,有其他方案未能覆盖的考量,但想要构建统一、全面、完善的人工智能准则既是难以实现的,又是没有必要的——难以实现的原因是人工智能这门科学技术本身、人工智能准则、伦理的内涵和外延都是在不断完善的;没有必要的原因是每个国家、组织与机构的准则提案都结合自身实际情况,有组织目标、环境、文化、伦理传统相关的特殊考虑。

笔者认为,真正有价值的做法是认可每一个国家、机构提案在一定范围的意义,而为了更好地实现人工智能的全球治理,重点将不是统一人工智能的准则,而是以更和谐互补的方式链接不同的人工智能准则提案,使其在局部(如国家、组织机构等)有序与一致,世界的全局范围内不同的准则提案仍可交互和协商,最终实现和谐互补、优化共生。

2 调和政府与企业在准则制定上的现存差异

在对不同人工智能准则提案的量化分析中可以看到:相对而言,各国政府对人工智能的潜在风险与安全高度重视,但企业的重视程度相对薄弱。这反映了在企业从事人工智能创新过程中对于潜在风险与安全隐患可能估计不足。例如,在对人工智能风险的评估中,部分企业认为如果总体可能的利益远远超过可预见的风险和不利因素,则可以进行相关探索。但从学术的视角看,一方面,依据潜在利益与风险之间的量化差异决定企业是否采取相关行动本身就是一个危险的视角;另一方面,如果没有站在全社会的全局视角进行综合研判与分析,仅从企业自身视角进行预测与判断,则很有可能因为局限的思考与行动对社会造成潜在的巨大危害。

因此,应当看到政府重视与企业相对不足的考虑之间的差距,并采取引导、监管、建立人工智能产品与服务的风险与安全评估体系等有效措施弥补政府期望与企业实践之间的鸿沟。

同时,一些准则在相关技术途径可能引发风险的方面可能估计不足。例如某些提案中对通用人工智能(各个认知功能达到人类水平,简称AGI)、超级智能(所有认知功能超过人类水平,简称ASI)涉及很有限。

这方面学术机构的讨论和相关研究应当为政府决策提供有力支撑,例如阿西洛玛人工智能准则中提出超级智能以全人类受益为标准来发展,剑桥大学近期提出并正在进行一项名为“通用人工智能的实现途径及潜在风险”的研究。

3 兼顾专用人工智能和通用人工智能的发展

在人工智能顶层设计当中,是否发展通用人工智能,还是仅限于发展领域特定的专用人工智能?这是诸多人工智能准则提案中主要存在的巨大分歧。

例如,欧盟人工智能研究实验室联合会(CLAIRE)就明确提出应限制人类水平智能、通用智能以及超级智能的发展。德国的人工智能计划也集中于发展专用人工智能。而阿西洛玛人工智能准则、OpenAI等机构提出的人工智能准则中则明确提出应当对更安全的通用人工智能进行充分考虑与发展。

事实上发展专用智能并不一定能够完全避免风险,因为专用智能系统在应用中很可能遇到不可预期的场景, 具备一定的通用能力反而有可能提升智能系统的鲁棒性和自适应能力。因此应针对智能的通用性这个主题链接不同国家、组织、机构的考虑,实现顶层设计的互补。

人类认知功能的多样性以及人工智能应用场景的复杂性使得很难对人工智能的风险、安全以及伦理体系进行绝对完善的建模,特别是涉及到人工智能产品、服务与具有不同文化背景的人类群体交互。任何国家都应制定适应本国的实际社会、科技、经济发展与文化需求的人工智能准则。但更关键的是需要全世界不同政府、学术组织、产业界深度交互与协同,对人工智能的发展进行对全社会有益的战略性设计,对其潜在的社会风险与伦理挑战进行系统性评估与预测,并构建世界范围内总体和谐、互为补充、优化共生的发展准则。

(作者系中国科学院自动化研究所类脑智能研究中心研究员、副主任)

爱思唯尔研究报告发现:

中国将在人工智能研究领域成为全球领导者

本报记者 齐 芳

日前,国际科技和医学信息分析公司爱思唯尔最近发布的一项研究报告《人工智能:知识的创造、转移与应用》。报告称,中国将在人工智能(AI)研究领域成为全球领导者,重要性日益凸显。同时,越来越多的研究人员正逐渐从学术界向产业界流动。

据介绍,报告中的数据不仅来自爱思唯尔旗下的Scopus数据库等科技数据库及平台,还参考了斯坦福大学人工智能指数报告、中国科学院自动化研究所数据集等一些数据信息。

报告显示在全球范围内,人工智能研究在过去五年(2013—2017)以每年接近13%的速度快速增长,而在2008年至2012年的五年间增速仅有不到5%。相比之下,过去五年(2013—2017)全球范围内所有学科领域的科研产出每年的增长仅为0.8%。

在检索分析了研究、教育、技术与媒体四大领域的共计60万份文档和700多个领域特定的关键词后,报告揭示出了人工智能关注的七个不同研究领域:搜索与优化,模糊系统,自然语言处理与知识表示,计算机视觉,机器学习与概率推理,规划与决策和神经网络。在以上研究领域,机器学习与概率推理、神经网络和计算机视觉的科研产出最高,增长速度也最快。

研究数据显示,在过去三年,中国学术界吸引人工智能人才的数量远远多于流失的数量,表明中国正在朝着人工智能研究领跑者的目标大步迈进。2004年,中国在人工智能领域的研究成果数量就已超过美国,如果保持当前的势头,中国有望在四年内赶超欧洲,成为全球人工智能研究成果最多的地区。

报告中有关不同区域的重要发现还包括:国际性流动与合作的模式表明,中国在人工智能领域的研究工作相对孤立。欧洲是人工智能科研产出规模最大、最多样化(人工智能相关研究领域的丰富性)的地区,与欧洲以外的国际合作比例及增长速度也很高。

研究还发现,尽管人工智能对社会发展的影响越来越深,媒体对人工智能在道德层面可能产生的影响也有关注,但是关于人工智能道德层面的学术研究还很有限。爱思唯尔分析服务高级副总裁玛丽亚·德·克莱恩认为:“新一代的技术(通常被统称为人工智能)的确非常重要,但人们似乎没有对其确切的定义达成共识。通过对人工智能科研表现的全面研究,我们希望明确该领域的动态、趋势和相关因素,并为其提供一些研究思路和见解。这份报告并不是对人工智能的总结陈词,而是帮助我们更好地进入人工智能时代并讨论日益发展的共生技术的开端。”

科技随笔

科学发展家

□王 姝

不久前发布的《2018中国科幻产业报告》中,有一组数据值得思考:2017年国内院线科幻电影市场总票房129.59亿元,其中国产科幻电影票房为13.17亿元。2018年上半年,国内科幻电影整体票房为95.06亿元,其中国产影片为8.9亿元,占比不足10%;

一方面,票房成绩显示出国内科幻电影市场巨大的需求和提升空间;另一方面,我们也不得不正视国产科幻电影票房成绩偏低的尴尬。如果将下半年陆续上映的电影票房统计进来,国产科幻电影在全国科幻电影票房中的占比有可能会进一步下降。

高票房成绩体现了国内观众对科幻影片的需求,而低票房占比则表明国产科幻电影尚不能满足观众的需求。如何提高国内科幻电影的质量,吸引更多的国内观众,是当下影视从业者迫切需要思考的问题之一。

对比发现,不同于好莱坞制作团队中频频出现的科学顾问,我国影视制作团队中难觅科学顾问的身影。科幻电影是一种基于想象力的创作,最重要的基础是从科学出发,与科学密切相关,科学内涵是科幻作品与神化、魔幻等作品最重要的差别。而国内科幻影视作品中的科学元素明显不足。在电影中,科学的价值不仅在于其中准确的科学知识,还在于制片人能否把科学作为一个创作手段,为作品增添精湛的视觉效果和扣人心弦的智力内容,从而增加影片可收获的潜在好评和票房,形成独特的良好口碑。这种口碑可以通过邀请科学顾问参与制作,并坚持科学真实性来实现。

影视中的合理性,取决于影视作品的画面和情节有科学基础,符合逻辑,能够自洽。随着我国教育普及比例的不断提高,当下观众具有的文化水平也在不断提升,电影作品中的画面和剧情如果太假太粗糙,很容易引起大众的抗拒,科学家在电影制作中的作用正在于此。科学顾问的目标是让影视制片人在他们的叙事、风格和受众的语境中实现科学的准确性,这种准确性,不同于科学家群体在科研探索中追求的准确性,而是一种相对的、更接近于大众可接受的“真实性”。

科学家是掌握最准确、最权威、最前沿科学信息的群体,他们的参与对于提升电影质量的意义已经被多次证实。科学家群体在帮助影视制片人创作合情合理、栩栩如生的故事方面显示出了惊人的价值。在漫威电影《雷神》的制作过程中,美国加州理工学院理论物理学家肖恩·卡洛尔(Sean Carroll)及其科学顾问团队,为女主角设计了“所有的魔法都是我们无法理解的科技”的台词,从而将原本属于北欧神话的雷神托尔顺利转变为来自更发达星球的外星人,为雷神成为漫威宇宙的超级英雄打通路径。同时,这些科学顾问还通过将女主角的职业从护士改为物理学家,帮助理顺剧情逻辑(一位护士深夜出现在沙漠里很奇怪,换成一位研究爱因斯坦-罗森桥的物理学家就很恰当了)。

科幻影视的制作人在推进影片制作的过程中,不妨考虑邀请接受过严格科学训练的科学家加入制作团队,感受一下他们解析细节、从复杂体系中找出相关联系的能力,在充实作品情节和确定结构框架时,借助科学家的力量,为提升影片质量打下更好的基础。

(作者为中国科普作家协会会员)

一场世界范围的铁路颠覆性技术革命

□ 沈志云

1964年日本建成东海道新干线,是世界第一条高速铁路。在既有米轨线以外,另行建立准轨全新高速铁路系统。采用整体道床,成倍加大最小曲线半径,成倍加大隧道截面面积。列车牵引采用动力分散模式,运营时速210公里。所有这些都是当时传统古老铁路所没有过的,在整体上与既有铁路根本不同,是典型的颠覆性技术创新。虽然起步速度偏低,但却立即风靡全世界,开辟了一个全球性的铁路颠覆性技术革命的新时代。

法国动作最快,20世纪80年代初建成从巴黎到里昂的高速铁路,运营时速270公里。后又建成大西洋线及欧洲之星,运营时速提高到300公里。最后建成的巴黎东线和地中海线,运营时速提高到320公里。可惜的是,因为过分强调降低成本,如仍采用传统碎石道床、传统的动力集中等,每车一个转向架也使轴重难于降低。严格说,难以达到颠覆性技术创新的高度。

德国最初的ICE1和ICE2也是采用动力集中,但他们很快发现动力集中对于提高速度不利,从ICE3开始,改为动力分散,运营时速也提高到300公里。对于碎石道床也发现精度不高,在高速下有碎石飞扬的毛病,故开始改用整体道床。可惜的是,他们对于必须形成独立的高铁网认识不足,高铁区段分散建设,只能与传统铁路线联运,而且客货混跑,不能建成独立的高铁网,难于进一步提高速度、发挥更大作用。

应当说法、德两国在研发高铁技术上都下了很大功夫,掌握了很多新的高铁技术,是这场世界高铁颠覆性技术革命的重要战场。但从总体来说,他们仍需继续努力。

我国从1978年开始进行高速铁路颠覆性技术的研发,晚来的中国高铁,却率先取得了这场技术革命的胜利。

1978开始的10年准备期间,通过改革开放,多渠道了解国外情况,分析总结各国经验教训,从理论上提升,形成高速列车大系统动力学,为系统仿真、系统优化、系统控制提供计算方法及软件。调查世界各国用于高速铁路技术的试验情况,为建设高速试验设备做好了充分的准备,1988年国家正式批准开始建设时速450公里的滚动振动试验台。20世纪90年代的十年探索,在原铁道部的统一领导下,各工厂纷纷研制高速列车,型号就有二十多种。这十年,初试锋芒,获取经验,培育人才,为高铁技术攻关打好基础。到了1998年,就进入实战的十年。内部开展采用轮轨技术还是磁浮技术的辩论,最终把轮轨高铁纳入国家规划,通过联合设计制造,创造中国品牌,达到了引进消化吸收再创新的最佳效果。

2008年8月1日,中国第一条京津高速铁路终于开通运营,最高运营时速达到350公里。2011年开始研制的复兴号,跑到了350公里的世界最高运营时速,而且在标准化、简统化等方面有很大提高。在高速路网建设方面,更加一日千里,从1.2万公里的四纵四横,发展到2.5万公里的八纵八横,联通所有50万以上人口的城市。

独立于既有线之外的、全部新建的中国高铁网,采用整体道床,轨道偏差标准达到毫米甚至零点几毫米级。最小曲线半径9000米(日本4500米)。隧道断面面积100平方米(日本64平方米)。高架及隧道占80%以上。以世界最高速度运营的复兴号高速列车,采用动力分散,超稳定(临界时速600公里以上),超平稳(可立硬币),超安全(每亿人次事故死亡率日本高铁0.9人,中国高铁0.2人),超环保(高架占地少),公交化运行,与城市交通连接。

四十年改革开放,造就了四十年我国铁路的颠覆性技术创新。半个多世纪以来,一直在进行的世界范围的铁路颠覆性技术革命,我们虽晚到十几年,却后来居上。高速铁路前景无限,让我们继续进行世界铁路颠覆性技术革命!

(作者系中国科学院院士、中国工程院院士、西南交通大学教授)

本报地址:北京市东城区珠市口东大街5号 邮政编码:100062 电话查号台:010-67078111 读者服务热线电话:010-67078442 定价每月30.00元 零售每份1.00元 京东工商广字第8063号