

从图灵奖得主“卷入”骂战看 AI “偏见”

■本报记者 卜叶

编者按

当我们还在谈论人类的刻板印象时,人工智能(AI)的“偏见”已经开始刺痛公众的心。近日,2018年图灵奖得主 Yann LeCun“卷入”了一场骂战。起因是 Yann 发表了一种 AI 算法 PULSE,能依据被识别图片生成一张真实却并不存在的人脸。“换脸”过程中,网友发现 PULSE“偏心”,生成的图像很单一,不能依据不同人群“量体裁衣”,导致有些人适配准确,有些人出现偏差的结果。Yann 在推特上指出,这是训练使用的数据集的局限性导致的,但公众并不买账,冲突愈演愈烈,而 Yann 因此退出推特。

偏见究竟是如何影响 AI 的?公众应如何对待不完美的研究?消除偏见最终是否可以实现?《中国科学报》特别邀请了业界多位专家就此进行了讨论。



讨论嘉宾

李甲 北京航空航天大学教授
张军平 复旦大学教授
雷震 中科院自动化所研究员
邓伟洪 北京邮电大学教授

邓伟洪

雷震

张军平

李甲

躲不过的偏见

《中国科学报》:AI 偏见是个常说常新的话题。在你们看来,究竟什么是 AI 偏见? AI 偏见有哪些表现形式?

邓伟洪:对于 AI 偏见,目前尚无明确定义。随着 AI 技术水平的提高和广泛应用,偏见等伦理和道德问题会凸显出来。世界领先的 AI 公司普遍受到该问题困扰,AI 领域是偏见的高危地带。

雷震:此前,亚马逊用 AI 代替人事部门筛选简历,后来发现筛选出来的简历均为男性,对女性求职者表现出偏见。这是因为 AI 根据过去简历的关键词来识别,而这些简历中,大部分求职者为男性,女性相关的数据太少,从而导致 AI 误以为没有这类关键词的女性简历不那么重要。

张军平:把人类、大熊猫识别成大猩猩,在阿拉斯加犬和哈士奇的识别中,只识别照片的背景,有雪地就是阿拉斯加犬,这些都是 AI 偏见的案例。

《中国科学报》:公众是否对 AI 偏见反应过激? AI 偏见会导致哪些问题?也是常见的现象。

李甲:AI 偏见是自然存在的问题,也是常见的现象。

雷震:一个算法模型在不同数据集、不同场景应用的结果有差别,这是正常的。

张军平:AI 偏见可能带来严重后果。比如,智能医疗影像诊断领域,就诊患者的诊断结果通过其他病人的数据训练而来。一旦出现偏差,这些患者的诊疗就会受影响。

《中国科学报》:AI 偏见能否避免?

邓伟洪:当前人脸识别主流的训练集是在互联网随机下载名人图片,这种貌似公平的随机下载,“无意中”把互联网的数据偏见传导给人脸识别算法。比如,目前不少训练集中的亚洲人脸数量不足 10%。

张军平:为了避免造成巨大偏见,建立公平的测试集是避免算法偏见的最有效手段。今年国际计算机视觉和模式识别会议(CVPR)上,我的研究小组发布了一个更均衡的数据集。目前,已有数十个国家和地区的研究小组申请使用。

雷震:从技术角度看,AI 能回答所有问题,但要考虑后果。一些可能触发“偏见”的研究要尽量避免,比如人脸预测犯罪倾向。

张军平:数据集不可能覆盖所有的可能性,算法偏见也许会持续存在。有些偏见不见得能马上被发现,目前依靠机构去纠偏并不现实。比较现实的做法是,科研人员在做情感识别、微表情识别、心理状态判断、AI 看相这类类型的研究时要谨慎,避免形成“以面积心”的偏见。

雷震:目前,主流的 AI 技术都是数据驱动的,多数偏见是由数据集的局限性导致的。此外,物理上的客观局限也会导致偏见。比如,儿童的相貌差距小,其人脸识别比成年人难,识别准确率低,这是客观原因导致的,并非对儿童有偏见。

李甲:通过反推,可以分析出 AI 偏见源自哪里。而往往,AI 偏见不是某一个原因导致的,而是多个问题,比如数据、技术、客观原因等共同作用的结果。

《中国科学报》:追本溯源,哪些原因导致了 AI 偏见?

邓伟洪:数据集和算法都会导致偏见。数据集偏见比较容易检测和消除,通过平衡和丰富数据可以有效消除。从技术角度,可以通过人工标注并统计训练数据的某个属性,从数量均衡的标准来确定数据集是否有偏差。算法偏差则更隐蔽,人们无法直接感知算法会偏爱某一组的用户对象,需要建立专门的测试集来检测,并且要面向公平的数据集来设计和优化模型,才有可能检测和避免算法偏见。

雷震:目前,主流的 AI 技术都是数据驱动的,多数偏见是由数据集的局限性导致的。此外,物理上的客观局限也会导致偏见。比如,儿童的相貌差距小,其人脸识别比成年人难,识别准确率低,这是客观原因导致的,并非对儿童有偏见。

李甲:通过反推,可以分析出 AI 偏见源自哪里。而往往,AI 偏见不是某一个原因导致的,而是多个问题,比如数据、技术、客观原因等共同作用的结果。

《中国科学报》:能否通过优化算法或更新数据集等技术手段消除 AI 偏见?

邓伟洪:规避偏见是一个系统方案,需要在数据集和算法上同时着手。公平的数据集可以消除一部分偏见,但不能完全消除。例如,北京邮电大学建立了统一的人学考试和录取标准,但也需要某些特殊群体加进来实现更广泛意义的“公平”。学术界已有的纠偏机制人为地去去除体现偏差属性的特征,这些方法往往会牺牲系统的整体性能。

雷震:对于偏见,可用技术手段纠偏,但不能完全解决。数据总有一定的倾向性,收集数据受时空限制难免有偏差,难以均衡地、全面地涵盖所有类型的数据。目前,数据集公布的信息较少,一般只公布男女比例、大致年龄范围、数据如何采集等基本信息。

李甲:通常,一个数据集中,有的类型的图片多,有的少。数量多的类型识别率高,数量少的识别率低。通过对数据集打一个补丁,优化数据集可纠偏,并且技术难度不大。此外,学术界对小样本群体如何提高识别率,已经展开前沿研究。

张军平:只能尽可能减少 AI 偏见,但不能完全杜绝,“例外”总是存在。如

公平的数据集可以消除一部分偏见,但不能完全消除。例如,北京邮电大学建立了统一的人学考试和录取标准,但也需要某些特殊群体加进来实现更广泛意义的“公平”。学术界已有的纠偏机制人为地去去除体现偏差属性的特征,这些方法往往会牺牲系统的整体性能。

雷震:对于偏见,可用技术手段纠偏,但不能完全解决。数据总有一定的倾向性,收集数据受时空限制难免有偏差,难以均衡地、全面地涵盖所有类型的数据。目前,数据集公布的信息较少,一般只公布男女比例、大致年龄范围、数据如何采集等基本信息。

李甲:通常,一个数据集中,有的类型的图片多,有的少。数量多的类型识别率高,数量少的识别率低。通过对数据集打一个补丁,优化数据集可纠偏,并且技术难度不大。此外,学术界对小样本群体如何提高识别率,已经展开前沿研究。

张军平:只能尽可能减少 AI 偏见,但不能完全杜绝,“例外”总是存在。如

果我们想穷尽偏见,那可能会触发组合爆炸问题。

让 AI 研究“喘口气”

《中国科学报》:有一种观点认为“AI 追求的不应是无偏见,而是透明度”,用透明度换取信任,对此你们怎么看?

邓伟洪:透明度可以解释偏见的来源,方便与公众的沟通,但不能解决问题。我们需要在伦理准则的指导下,建立由公平的标准、训练数据和算法模型等构成的多层次研究体系,完善地解决问题。

雷震:学术论文旨在服务学术交流。为了确保结果的可重复性,一般论文中的数据是开源的,实现方式是透明的,代码是公开的,业内人员能够清楚了解数据集的缺陷。从这个角度看,透明是好事,但透明到什么程度,是不是所有问题都告诉公众,全部告知是否会引发其他问题,也是未知的。

张军平:科研人员不是先知,在研究阶段不一定能洞见所有问题。数据集、算法等存在哪些问题,如果研究前就清晰了解,那研究也就没什么意义了。如果透明度能定义,那科研就不是科研了。科研本身就有很多未知因素掺杂其中。

《中国科学报》:有科研人员认为“不同于应用,研究都是不完美的”“AI 的输出结果都是带有不确定性的”“不要让研究者们每天活如履薄冰”,你们希望公众怎样看待 AI 研究?

邓伟洪:AI 是一个方兴未艾的研究方向,希望公众对 AI 研究持一定的包容态度。在当前情况下,我们可以通过调节系统参数来平衡准确性和公平性。

雷震:AI 获取了人类社会的知识和数据,人会犯的错误,AI 也会犯。不可否认,纠偏机制是必需的,科研人员要及时改正偏见问题。同时,也要尽量把 AI 看成人类,让它适当“喘口气”。

李甲:偏见的消除是逐步的,需要时间。但是,公众希望时时刻刻得到四平八稳的结果,不希望偏见发生,这时候就需要媒体做好“传声筒”,一方面告诉公众,科技一直在进步;另一方面告诉科研人员公众的诉求。

浅谈新时期人工智能如何赋能国家社会经济发展

■孙守胜 董晓英

自新冠肺炎疫情暴发以来,从湖北武汉到整个中国再到整个世界,社会经济发展面临巨大挑战。如何稳定社会形势,找准经济发展的新路径,是摆在各国面前的难题。本文将疫情为背景,浅谈中国产业发展的新挑战与新机遇,分析人工智能在新时期将如何赋能国家与社会发展。

一、发生在高度全球化世界中的“大疫情”

1. 疫情对世界经济产生“百年难遇”的影响
此次疫情正在全球范围内产生损失惨重、历史罕见的巨大影响。1997 年的亚洲金融危机波及最严重的地区仅为泰国、韩国、日本等亚洲国家,中世纪的黑死病与 2003 年的 SARS 也具有较为明显的区域性特征。但截至 7 月 7 日,中国的新冠肺炎确诊病例确诊感染超过 8 万例,世界累计确诊已逾 1150 万例,呈全球蔓延趋势。从全球经济数据来看,此次疫情更是导致世界各国经济元气大伤。根据国际货币基金组织(IMF)预测,今年全球经济总量将萎缩 3%^[1]。作为对比,这一数据在金融危机后的 2009 年也仅为 0.7%^[2]。而目前尚不明朗的疫情发展趋势与各国随之而来的政策调整更意味着疫情影响的影响不只发生在当下,更对未来提出巨大挑战。

2. 内外冲击下的中国经济

我国第一季度国内生产总值同比下降 6.8%,多项重要经济指标均出现负增长,社会经济在内外冲击之下面临困境。

一方面疫情冲击国内生产和市场。隔离与交通中断限制了人们的生产消费活动。工人因疫情延迟返工,人员密集作业环境复工难度大,产业链上下游协同受到阻碍^[3]。社会活动停摆、失业率上升等因素也导致居民消费水平的下滑。

同时,全球化产业链中断带来外部冲击。在全球范围内扩散蔓延的疫情,使全球产业供应链承受着巨大的压力。根据海关数据,今年前两个月我国货物贸易进出口总值 4.12 万亿元,比去年同期下降 9.6%^[4]。一项针对国内外贸易企业的调查显示,未受疫情明显影响的企业不足 2%^[5]。

目前,国家主管部门及权威专业人士大都认为国内经济复苏趋势较为积极。据工信部调查,5 月全国中小企业的复工率已达 91%,复工复产有序进行,成效显著。而根据国家统计局公布数据,2020 年 1 到 5 月份中国 PMI 指数分别为 50.0%、35.7%、52.0%、50.8%、50.6%,显示制造业经济逐步复苏回到扩张期,经济趋势整体稳定向好。当然,区别于非常时期经济一个季度内快速回温的 V 型曲线^[6],本次新冠疫情传播的时间差导致疫情产生的负面影响相对分散,经济复苏所需过程也就相对漫长,预期呈现 U 型曲线^[7]。如何更进一步加快经济稳定复苏的脚步,将 U 字的底部收窄,也是当前的棘手挑战。

二、中国产业发展面临的机遇与挑战

1. 逆全球化的低可能性

在这场作用范围甚广、影响力度极强的疫情之中,国际经济形势似乎呈现出逆全球化表现。然而实际上,经济全球化是各国各地区间优势互补、实现效益最大化的市场自然选择,是难以通过国家政策完全扭转的发展趋势,在相当长的一段时间内仍会是世界经济发展的主流。正是因为在全球化的背景下充分运用分布广泛的资源与技术,产业发展才得以拥有现今的高度。打破成熟的跨国产业链并在一国之内重建,难以实现。这种“逆全球化”现象并不符合国际经济发展的规律,经济全球化仍会在一定时间内为我国产业提供机遇。

2. 中国当下产业发展依旧保有优势

经过多年积累,我国产业发展保持着相当的国际竞争力。一方面,中国当前仍占据着世界工厂的地位,制造业门类齐全,工业体系完整,是全世界唯一拥有联合国产业分类当中全部工业门类的国家,目前有 220 多种主要工业产品的产量居全球第一^[8]。另一方面,依靠着广阔的内部市场,我国一定时间内仍将保持稳定的经济增长。多项外资企业调查显示,中国市场仍然备受外国投资者重视^[9]。

3. 产业发展背后存在不可忽视的瓶颈问题

尽管当下全球经济环境与产业积累呈现积极趋势,我国产业发展也同样存在着不可忽视的问题与挑战。首先,我国核心技术依然存在受制于他的现象。在全球化的产业链中,整体仍处于产业链、价值链中低端,核心技术突破缓慢,难以支撑产业高端化发展^[10]。第二,劳动力成本优势逐渐消失。由于国内劳动力成本升高,一部分低端制造业已经向东南亚等国转移,我国制造业的劳动力成本优势正不断减弱。在疫情的冲击下,以上制约产业发展的瓶颈问题更加凸显,如何应对挑战,更好地化优势为产业发展的动力是当前推动社会经济发展亟须解决的难题,科技创新与产业升级迫在眉睫。

三、科技创新赋能:人工智能助力社会经济高质量发展

1. 政策指明产业发展的新方向

梳理国家产业发展的中长期规划。从《中国制造 2025》强调信息化与工业化深度融合,到 2017 年《新一代人工智能发展规划》对于抢抓人工智能发展机遇、构筑先发优势的要求,科技创新在产业发展中的重要地位毋庸置疑。而作为第四次工业革命的技术基石,人工智能更将为产业升级与经济高效发展注入强大动力。

在 2018 年的中央经济工作会议中,“新型基础设施建设”(以下简称新基建)的定义被第一次正式提出,5G、人工智能、工业互联网、物联网成为基础设施建设新的重点发展方向。其中,人工智能作为新基建的主要领域将带来巨

大的经济与社会效应。大力推进人工智能基础设施建设与实际应用,可以更好地服务经济和社会^[11]。为新一轮产业变革提供强大的动力。基于 5G 的通信网络基础设施,为人工智能的高效应用提供了必要条件,云计算、区块链等也必须基于人工智能的升级发展才能真正得以实现,互为依托、互为应用,显示着人工智能广阔的发展空间与发展实力。

2. 融合人工智能的产业对经济有重大催化作用

结合新基建政策内涵,融合基础设施是新基建的发展方向,也是人工智能等领域科技创新成果落到实处的关键所在。在与传统产业的融合中,人工智能将赋能生产,打造新的经济发展点,为经济发展提供强大动能。

(1) 人工智能为传统产业赋能

人工智能作为赋能技术,可以放大资本和劳动力当前的能力^[12],将其与制造、医疗、教育等领域进行融合,可为传统行业带去新的生命力,优化资源配置,提高生产效率。根据《新一代人工智能发展规划》,2030 年人工智能核心产业规模应超过 1 万亿元,带动相关产业规模超过 10 万亿元,足见人工智能对产业发展的催化作用。充分应用人工智能研究成果,推动研发、生产、运营向数字化、网络化、智能化转型,可以使我国工业体系完备、产业集群全面等优势得到充分发挥,大幅提升企业创新能力和集约化发展水平,通过“机器化换人、智能化减人”,提高生产效能,增强制造业实力与发展稳定性。以本次疫情中企业的表现为例,数字化程度越高的企业越能够将疫情的冲击降到最低^[13],人工智能在传统生产领域的融合应用可加快传统产业的数字化、智能化,降低企业对密集劳动力的依赖,同时也能增强企业应对全球产业链复杂变化的实力。

(2) 形成新的市场需求

随着人工智能的发展,新的市场业态与需求也在逐渐形成。人工智能技术的落地催生了新的经营方式与新的智能化产品,其应用又会

改变人们的生活方式,进一步影响、塑造新的市场需求。疫情期间,线上办公、远程会议等新的应用一跃成为最热门的消费点,降低人员接触的要求也带动了人工智能机器人在医院等场景的使用。根据我国第一季度 GDP 数据,在其他领域均呈下滑趋势时,信息传输、软件和信息技术服务业则保持增长态势,比上年同期增长 13.2%,足见信息技术行业巨大的市场需求与增长潜力。同时,当前流行的线上消费形式,诸如线上教育、视频会议等,正在或者已经培育了人们新的消费习惯,这些习惯并不会随着疫情的结束而消失,反而会进一步塑造人们的生活方式与生活习惯,相对应的消费需求在更长时间内也会得到延续,并带动新的产业发展。

3. 融合人工智能的产业能够推进社会治理,改善民生

人工智能产业不仅能够成为推动经济转型发展的加速器,也可以在应用于社会生活的过程中服务于民生,为创新社会治理方式、提高社会治理效率与水平提供可靠支撑,有力推动社会治理的智能化转变。

(1) 人工智能应用促进社会治理方式更新换代

随着数字化、智能化平台的构建,人工智能的应用可以协助政府机构加快创新治理模式,在政府工作中发挥越来越重要的作用。一方面,人工智能辅助政府管理方式创新升级。其应用可以为政府工作提供强大的分析支持,借助国内广阔的数据资源,生成可信的动态数据检测报告,为政策制定提供科学的依据。未来,社会治理也将形成经验与创新良性循环,实现社会治理结构的自主优化,在社会民生中发挥更大的作用。另一方面,人工智能使社会治理更加开放透明。智能化平台可以加速信息的社会共享^[14],为社会大众参与社会治理提供更多的可能性。让更多主体参与到社会治理中来,不断丰富社会治理内涵,并让大众的声音在政策制定中听得见、有回响,实现共建共治共享的社会治理格局。

(2) 人工智能技术助力社会治理资源整合,提升治理水平

通过信息整合与智能化平台的构建,人工智能技术可以联通更多行业,快速获取实用信息,对复杂多样的数据资源进行有效整合,避免政府工作的重复与资源的浪费,进而提升社会治理效率。同时,人工智能技术能够提升社会治理水平与效果,基于大数据平台掌握社会、行业发展的趋势,提高预判能力,使政策制定更加及时有效,社会治理更加积极主动。以疫情期间的的大数据平台为例,人工智能在预判疫情走势、公开疫情数据等方面均发挥了不可忽视的作用。

4. 规范人工智能产业发展,建立科学的发展体系

人工智能的发展是机遇也是挑战。在促进经济发展、便利社会治理的同时,新技术的应用与大量数据的挖掘也带来了一系列监管与治理上的难题。

为规避人工智能广泛运用可能带来的风险,首先我们应不断完善相关制度和规划,建立强大的制度保障与制约机制,集结跨学科智慧搭建制度框架和科学发展规划体系,加强对人工智能文化层面的学习研究,赋予其健康发展的“软”力量。第二,在技术层面,需加强对数据安全与保障工作的,并严把技术关,提升技术的利用效率与应用安全。同时,应加强国际层面的人工智能交流合作,鼓励学界、产业界积极参与国际规则制定,在未来人工智能发展中发挥积极的大国作用。

疫情为社会发展带来冲击,也必然会加速科技创新以带动社会升级发展。面对我国产业发展的优势与瓶颈,人工智能毋庸置疑将成为推动社会进步、服务社会民生的有效推动力。目前,我国人工智能产业发展仍然存在技术成果转化跟不上市场需求增长的问题,人工智能依托的数据共享依然需要国家整合,如何更好促进人工智能与传统产业融合发展,寻求新时期技术成果转化的创新范式,是国家社会经济面临的重要现实挑战与迫切任务。

(作者单位:中国科学院自动化研究所)

参考文献:

[1] 国际货币基金组织(IMF).《世界经济展望》[R].华盛顿:2020.04.
[2] 江玮.专访特里谢:欧洲不差钱,但当前危机比“大萧条”更棘手.《财经》,2020-05-03.http://yuanchuang.caijing.com.cn/2020/05/03/4662028.shtml
[3] 洪卫.疫情冲击对制造业供应链的影响分析及“后疫情时代”的政策取向[J].西南金融:1-10[2020-04-23 11:19].
[4] 娜娜.全球疫情蔓延对外贸影响不可小觑.中国经济时报,2020-03-09.http://lib.cet.com.cn/paper/szb_con/513041.htm
[5] 李前.外贸企业及外贸供应链受疫情影响

调查报告.进出口经理人,2020-04-10.http://www.tradetree.cn/content/781120.htm
[6] 陆正明.我国经济已摆脱非典影响 三季度 GDP 增百分之九点一,预计全年增百分之八点五.东方网,2003-10-18.http://news.sina.com.cn/c/2003-10-18/0912941992s.shtml
[7] 全球经济将迎来哪种复苏? V 型? U 型?.长江商学院,2020-5-13.
[8] 降彦彰.中国经济如何逆势复苏? 专家建议全年 GDP 增速目标 3% [N]. 财经, 2020-04-17(http://news.caijingmobile.com/article/detail/415291).
[9] 李雁辛.工信部:我国工业实现了历史性跨越[N].上海证券报,2019-09-21(要闻).
[10] 王林,许亚杰.世界需要苏俄的“世界

工厂”[J].公共世界,2020,(08):49-50.
[11] 涂圣伟.突破产业高质量发展的内在瓶颈[N].大众日报,2018-10-10
[12] 蔡自兴.中国应加快发展人工智能“新基建”.中国新闻网, http://www.chinanews.com/cj/2020/04-09/9152020.shtml
[13] Accenture. Artificial Intelligence Is The Future of Growth [R]. https://www.accenture.com/us-en/insight-artificial-intelligence-future-growth
[14] 腾讯研究院 陈维宣 吴绪亮.新型基建促进经济发展的作用与路径[N].2020-05-08(011).
[15] 孙照.人工智能开启社会治理智能化转向[N].2020-05-27(006).