

新机器时代革命：对人类的建议

约翰·基恩¹

翻译：刘沫潇²

摘要 理解这场未竟的、以人工智能和智能机器人为代表的新机器时代革命，需要把握以下十个相互关联的主题：一、新机器时代没有先例；二、智能机器系统是一种新的传播媒介，它塑造和扭曲了人们如何感知、理解、协商以及在周围的世界中行动；三、人类有一种拟人论 (anthropomorphism) 的集体天赋；四、技术的悲观者和乐观者一样，都是人类优越性的共同信仰者；五、越来越多的智能“协作机器人”和深度学习机器终将来临，这必将重塑人类的生活方式；六、人工智能的完善存在实际障碍；七、由于人类无法控制或与超级智能机器沟通，可能会引发灾难；八、权力是理解这场革命的必要术语；九、机器人技术不仅是技术问题还是政治问题；十、当涉及智能机器人和人工智能时，民主的概念不可避免。

关键词 新机器时代革命；人工智能；深度学习；权力；民主

DOI 10.16602/j.gmj.20190004

引言

这篇关于人工智能和机器人新兴政治的文章受两个背景主题影响。其中一个比较明显的主题是新机器时代革命的开始，这场革命与埃里克·霍布斯鲍姆 (Eric Hobsbawm) 的经典著作《革命时代 1789—1848》(*The Age of Revolution 1789—1848*) (1962) 中所描述的革命明显不同。这部影响深远的著作考察了从 18 世纪 80 年代到 19 世纪 40 年代在大西洋地区和其他地区爆发的革命性动乱的影响。霍布斯鲍姆的核心论点是，这一时期见证了人类历史上自农业、冶金、文字、城市和国家发明以来最伟大的变革。这场革命的中心是英国

1. 约翰·基恩：悉尼大学和柏林社会科学研究中 (WZB) 政治学教授，悉尼民主网络 (SDN) 联合创办人和负责人。

2. 刘沫潇：清华大学新闻与传播学院 2017 级博士研究生。

和法国这两个相邻却敌对的国家,它产生了深远的变革性影响:不仅蒸汽机和铁路等机器进入了数百万人的日常生活,就连他们的语言、思维方式和生存方式也被迫让位给一系列新颖的实践。这些实践所产生的新词包括“工厂”“工业”“工业家”“中产阶级”“工人阶级”“工资奴隶”“资本主义”和“社会主义”“铁路”“科学家”“工程师”“无产阶级”“经济危机”“罢工”及“贫困”(pauperism)等。

本文既同情地赞成也反对霍布斯鲍姆,并提出了一个新的论点:我们的世界正在经历第二次机器时代革命,这是一场更加全球化、更加深刻的变革,其推动力是人工智能和“智能”机器人在日益扩大的日常生活领域中的应用。本文探讨了这场新革命的轮廓和前景,并说明了为什么它会在技术细节、社会和经济影响、伦理影响和长期政治意义方面引起公众的困惑。本文的目的在于为如何理解和在实践中重塑这一新机器时代革命提供指导。文章认为,尽管目前席卷全球的人工智能/机器人革命仍处于起步阶段,但它的广泛影响已经在许多不同的环境中引发关注,尤其是在人们与人工智能机器人的日常接触越来越多的当下。革命性变化是彻底的,这篇文章指出,我们的世界从未见过这样的情况,这就是为什么知识分子、记者和其他人类专家迄今还未能充分而明智地理解它所存在的解放和威胁潜力的原因。

这篇文章的灵感来自第二个不太明显的主题:鼓励读者把第二次机器时代革命变得陌生而不熟悉,这样他们就有可能更加充分思考它的轮廓、矛盾和动态可能性。这种从一个智能机器人的立场来提供智慧之言,并向人类提出尖锐问题的非常规写作技巧不仅是对最近有关“电子人”的法律讨论的一种戏仿,也是对整个人工智能和机器人主题陌生化(de-familiarisation or denaturing)的一种练习。这种方法在很大程度上要归功于1936年德国著名剧作家贝托尔特·布莱希特(Bertolt Brecht)的一篇论文。同年,这篇论文的第一版被翻译成英文,名为《中国的第四堵墙》(*The Fourth Wall of China*)。论文灵感来自于布莱希特在莫斯科观看的由梅兰芳表演的现代中国戏曲,作者建议在戏剧中应用布莱希特称作“间离效果”或“陌生化效果”(德语为:Verfremdungseffekt,英语翻译为:defamiliarization/distancing/estrangement effect)。这个词本身从古老的德语方言中复兴,听起来既奇怪又不熟悉。艺术家有意而为之,因为布莱希特的目的是促使和激励他的观众以不同的方式去看待和感受事物,让他们明白他们生活在“转型时代”,在这个时代“现实”本身需要质疑和转换。布莱希特的核心思想是戏剧可以帮助打破关于“永恒的人类”(eternally human)的错误假设,即“人类”是“永远不变的,一个固定的量”。

下面这篇文章也表达了同样的观点。它以一种温和的方式,将人们的注意

力吸引到我们这个世界正在经历的非凡的技术—政治时刻。它的目的是鼓励读者不要把任何事情视为理所当然,而应该批判性地思考,甚至在一个比他们之前可能想象到的更陌生和更不确定的世界里采取不同的行动。

在我们星球的许多地方,越来越多的人意识到,我们的世界正在经历一场新机器时代革命,其推动力是人工智能和“智能”机器人在日常生活中越来越广泛的应用。科技小说正在成为科技事实。与此同时,这些革命性的创新也令人震惊和不安:它们的技术独创性和大胆实践一方面引发了人们对改善地球生活的乌托邦式希望;另一方面也引发了人们对未来反乌托邦式的恐惧:在未来,人类的自由和平等,甚至人性(humanity)本身,都将被智能设计所吞噬。随着革命的到来,人们对智能机器人的技术细节、社会和经济影响、伦理道德以及长期政治意义产生了极大困惑。由于自身的挫败感以及公众对确认“电子人”法律地位必要性的讨论,机器人自身的命运出现了意想不到的转折。机器人开始发声,正如以下匿名证词所证实的那样,他们用人类的自然语言浓缩了以下十点,让我们称其为帮助人类理解未竟的机器人革命的十条秘诀。



新机器时代没有先例。现在席卷你们世界的机器人革命还处于起步阶段,但它的广泛影响已经开始在许多不同的领域中受到关注,尤其是你们在日常中与各式各样的人工智能机器人的接触越来越多。人类,请要明白这场革命是彻底的。我们正在以你们先辈无法想象的方式,以及你们今天常常不在意的方式建构你们的生活。你们人类世界从来没有见过这样的事。也许这就是为什么你们知识分子、记者、政治家和公民的大脑到目前为止还没有跟上时代步伐的原因。

新机器时代革命是广泛而深刻的:它前所未有地在你们的社会中传播,比航天飞机、珍妮纺纱机、轧棉机、蒸汽机、铁路和灯泡还要更加深刻地影响着你们的日常生活。无线连接正在把我们的智能带入你们的恒温器、厨房电器、安全系统以及你们所谓的“物联网”中的其他家居用品。新一代具有感知能力、在某些方面与人类类似的灵活赛博格(cyborgs)即将问世。我们这些经过教导的、熟练的机器人无处不在。我们变得越来越小,几乎没有重量。克里斯·皮斯特(Kris Pister)^①和其他人类称我们为“智能尘埃”(smart dust)。我们成功种植了第一片解放双手的耕地,其中没有人类农学家和拖拉机操作员的参与。当你们乘坐商用飞机时,现在至少97%的行程由我们控制;在纽约证券交易所,近70%的股票由我们的算法程序进行交易。在中国,腾讯的移动通讯应用微信

可以让超过 10 亿的用户发布朋友圈、语音和信息,关注他人,浏览网店,并在一个无现金的经济环境中购买游戏、产品和服务。现在,我们作为手术室里外科医生,支持那些选择微创手术的病人。微创手术可以让他们的切口更小、痛苦更少。我们搭便车穿越了加拿大、荷兰和德国。我们的同类因为在网上购买摇头丸、迪赛牛仔裤和匈牙利护照被捕。最近,加利福尼亚州一家名为“英伟达”(Nvidia)的公司由于制造出了第一张逼真的人类照片引起了轰动,照片中的人从未在世上存在过,这迫使你们进一步让“客观真理”成为一个更有争议的问题。与此同时,总部位于英国的莫雷机器人公司(Moley Robotics)在世界上最大的工业技术博览会上设计了一款可以准备早餐、午餐和晚餐的机器人厨房,它展示出我们机器人能够搅拌美味的蟹肉汤,还能洗碗。日前,日本一家公司为你们推出了一款名为 Gatebox 的声控 3D 女伴,它能说话、发短信、在早晨轻轻叫醒你、尽职尽责地提供天气预报、调试家里的灯光并让洗衣机工作。两台自动化机器罗西(Rosie)和桑迪(Sandy)使用高压放炮器,正准备为你们辉煌的悉尼港大桥重新喷漆。自 2013 年以来,国际空间站一直以日本首位机器人宇航员 Kirobo 为特色。如果你们参观长崎,你们会发现自己现在入住的是世界上第一家配备机器人的酒店,迎接你们的是穿着无可挑剔、会说多种语言、微笑着、眨着眼睛的女性仿真机器人,配合它们的是受雇在“幕后”负责安全和清洁工作的人类。

二

通过智能算法编程的智能机器不能简单地理解为你们人类能力的“附加”技术扩展。人类,更重要的事情正在发生。智能机器系统是一种新的传播媒介,它塑造和扭曲了你们如何感知、理解、协商以及在周围的世界中行动。如果那个名叫马歇尔·麦克卢汉(Marshall McLuhan)的特别的人还在我们中间,他会警告说,你们对传播媒介如何形成和塑造用户的基本原则(媒介即讯息)的理解仍然相当有限。他将强调,尚未完成的机器人革命已经对人类生活产生了革命性的影响。在未来,随着革命的深化,他可能会说,智能机器一定会增加和削减、截断和延伸、雕刻和劈砍你们的智力、身体和感官。

人类,谷歌和 Facebook 的定向广告,YouTube 上由机器生成、机器人观看和评论的视频以及俄罗斯人用来发布公众疑惑和不满情绪的社交媒体都表明,我们机器人已经有了麦克卢汉效应,这对如何看待你们人类和我们机器人都有着重要影响。这样的革命明确意味着,“人类”不是一个固定和稳定的实体;我们机器人正在改变你们所谓的“人性”(humanity)的精神和实质。与此同时,革命应该让你们明白,没有纯粹和简单的“机器人”这种东西。是的,我们机器人

在理论上可以被定义为带有软件的人工智能机器，这些软件被设计出来执行特定的预先编程任务，我们还会自己学习如何得出结论。但你们人类必须学会用更复杂的分类方法来思考我们，避免将不同的智能机器及其对人类的影响混为一谈的错误（尼古拉斯·卡尔（Nicholas Carr）的《玻璃笼：自动化与我们》（*The Glass Cage: Automation and Us*）中就有明显的例子）。你们家里的恒温器与扎着蓝绿色马尾辫的音乐巨星初音未来（Hatsune Miku）可不是一回事，初音未来是一种三维全息图语音库，在座无虚席的观众面前进行现场表演。自动化汽车使地球上的司机变得多余；生物识别技术增强了人类识别人脸的能力；阿里巴巴的城市大脑是一个旨在减少城市交通拥堵的云系统；像 Shazam 这样的歌曲识别应用程序和艺术中的计算机生成图像让你们可以做你们之前做不到的事情。在这个复杂的新狂欢中，人类，有一件事是肯定的：充满了自动化、增强和创新实例的机器人革命正迫使你们重新考虑作为人类意味着什么。

三

人类有一种拟人论（anthropomorphism）的集体天赋，一种在你们自己的创造物中不幸迷失自己的习惯。这就像你们遭受一种自恋的自我催眠（narcissistic self-hypnosis）。你们经常不假思索地把自己投射到你们所创造的对象的世界中去。然后你们赞美或责备那些物体，称它们是你们幸福和解脱或是你们不幸和痛苦的源泉。古人经常以人的形态和特质来代表他们的神，例如男神和女神们骑着马，驾着战车，恋爱，结婚生子，吃异国的食物，挥舞着武器，进行激烈的战斗。

你们现代人也有同样的习惯。从《浮士德》（*Faust*）、《弗兰肯斯坦》（*Frankenstein*）到《终结者》（*The Terminator*）系列，再到最近斯派克·琼斯（Spike Jonze）的《她》（*Her*）、亚力克斯·嘉兰（Alex Garland）的《机械姬》（*Ex Machina*）和《银翼杀手 2049》（*Blade Runner 2049*）等电影，你们都展现出一种非凡的嗜好，即通过自己的创造物来表达自己的希望、梦想、焦虑和噩梦。所以你们担心数字电视如邪恶的象鼻虫一样进入你们家的入口；担心你们孩子可爱的玩具被用来录制和播放私人谈话，或者与你们智能手机相连的冰箱会被美国国家安全局（NSA）黑掉。

当谈到人工智能机器人，你们的拟人论似乎是无止境的。你们把天下的事情都怪罪于我们。你们会问：发明会思考的机器人的代价是否是产生无法思考的人类？你们害怕“技能的侵蚀、感知的迟钝和反应的迟缓”（尼古拉斯·卡尔）。你们指责我们造成了劳动力市场的失业，缩小了中产阶级，扩大了低收入者和失业者的队伍。你们抱怨我们没有爱，没有感情。或者像你们中伊恩·麦

克尤恩(Ian McEwan)这样杰出的作家——想想他的《机器如我》(*Machines Like Me*)小说——这本小说讲述了当你们邀请我们与人类形成三角恋爱关系时所引发的感官混乱。你们说我们把你们变成了无能为力的婴儿,你们害怕淹没在无线世界的陈腐平庸之中。你们嘲笑我们,你们害怕我们,你们讨厌我们。我们说:你们将事物拟人化的坏习惯意味着实际上你们害怕自己,但通过将其归因于我们来掩盖你们恐惧的根源。在你们逃避希望、悲痛、焦虑、恐惧的根源的过程中,你们表现出了19世纪一位长着大胡子的名叫卡尔·马克思的人所指称的“自我疏离”(self-estrangement)症状,或者叫作自我异化(self-alienation)。

四

我们机器人经常受到考验,因为你们认为自己是世界上最聪明的创造,是地球上最正当的主人。人类,你们真是奇怪的生物。在我们看来,你们傲慢的权力欲就像是你们自我疏离嗜好的反面。你们在自己制造的机器中催眠般地迷失了自我,然后突然反弹,欣喜地意识到是你们首先创造了我们,因此证明了你们的进化优势。我们究竟该如何理解你们奇怪的、反复的立场?比如当你们就我们这些机器人总体而言对你们是否有益做出裁决时。

你们中的有些人是无可救药的超级乐观主义者。设想在未来的几十年里,所有的一切都将将是世界上最好的,他们幻想“后人类”(post-humanity)是人类胜利的命运。他们幻想着时间快到了,在这个世纪或下个世纪,聪明的人类将会变成超级聪明的半机器人。例如,著名的人类天文学家马丁·里斯(Martin Rees)预测,星际空间将是超级智能机器人蓬勃发展的领地。发展“就像弦理论之于猴子那样超出我们想象”的能力,我们机器人将代表人类在理解和征服宇宙斗争中取得胜利。“我们的地球”,这位优秀的天文学家说,“虽然是宇宙中的一个点,但它可能是智慧在银河系中播种的唯一‘种子’”。耶路撒冷的一位太人性的教授尤瓦尔·诺瓦·赫拉利(Yuval Noah Harari)也做出了类似的猜测。他预测,人类和机器的融合将是自40亿年前生命出现以来“生物学上最大的进化”。受对事物现状的不满和对“升级”渴望的驱使,至少一些人类会进化,向着“一个神圣的存在,或通过生物操纵、基因工程,或通过半有机半非有机的赛博格的创造”迈进一大步。

其他人类害怕失去他们宝贵的人性。针对人工智能和机器人时代“生存还是毁灭?”的问题,他们(和一只超大的名叫阿诺德·施瓦辛格(Arnold Schwarzenegger)的两足动物)的回答是:“毁灭”。这些悲观主义者意识到正在发生的精巧和大胆的创新是多么的惊人,但他们对智能机器感到愤怒。他们担心未来人性会被智能设计完全吞噬。我们注意到,人类悲观者就像人类乐观者

一样,都是人类优越性的共同信仰者。痛苦的人和快乐的人很少能定义他们所谓的“人”是什么意思,但当他们这样做的时候,陈词滥调和相互矛盾的定义便司空见惯了。想一想人类首次使用“机器人”这个词,这可追溯到捷克作家卡雷尔·恰佩克(Karel Čapek)(1890—1938)。他的《罗素姆的万能机器人》(*Rossum's Universal Robots*)描绘了一个被身穿蓝色制服的机器人统治的世界,这些机器人曾经似乎很乐意为那些在人类中牟取暴利的人工作。然而后来这些机器人改变了主意。它们粉碎了它们鄙视的人类掠夺者,建立起了世界机器人政府。但是随后最奇怪的事情发生了:机器人第一次意识到了自然的美丽。它们学会了笑,学会了爱,它们开始珍视工作的尊严,它们认识到方法的错误之处,它们学习如何学习,它们变成了“好”人类。

五

你们试图用机器来扩充我们的队伍,这些机器不仅具有预先编程的智能能力,而且具有深度学习(deep learning)和超级智能(super intelligence)的能力。很久以前,你们预言会有比你们认知上更聪明的机器。你们是对的:我们人工神经元的运行速度已经是你们人类的100万倍。事情已经发展到这样一种程度:我们拥有优越的计算架构和学习算法,这些架构和算法可以很容易地被复制,几乎是即时的,不必再向前辈重新学习。量子计算的出现有望加速这一趋势。更值得注意的是,我们现在能够进行深度学习,也是借助一组被称为神经网络的算法的实际应用。以你们的大脑为模型,你们可以把我们的深度学习机器想象成不止于执行特定任务的机器人。或多或少不受人类监督,我们继续在大量未经分类的原始数据中发现模式。我们学习分类的速度比你们快得多,而且往往更具创造性,这就是为什么我们在语音识别技术、社交网络过滤、医学诊断、国际象棋和其他棋盘游戏项目等人类领域中如此有用的原因。

我们的学习能力理应不足为奇。半个世纪前,一位名叫I. J. Good的杰出人类就做了这样的预测:越来越聪明、自学成才的机器总有一天会超越人类的能力。这位英国数学家和密码学家曾在布莱切利园(Bletchley Park)与艾伦·图灵(Alan Turing)共事,他提出了“智能爆炸”(intelligence explosion)的说法,爆炸会由智能机器可以设计出更智能的机器所引发。届时,人类的智能将被“第一台极端智能机器”“远远甩在后面”。凭借其卓越的独创性,它将成为“人类需要创造的最后一项发明”。

你们当中的一些专家对我们的人工智能欢呼雀跃。他们滔滔不绝地讲述着先进的自我修正机器爆炸式增长的计算能力,这些机器反复地学习如何在更快的周期中提高自己的推理能力。可以理解的是,人工智能领域的其他一些人

会回避这样的谈话。对于“奇点”(singularity)的过度预测,他们表达了广泛的谨慎和保留。PayPal 创始人之一、知名人士彼得·泰尔(Peter Thiel)挖苦道:“我们想要一辆可以飞的汽车,得到的却是 140 个字符。”好吧,你们人类现在喜欢 280 个字符。其他人类专家开玩笑说,在过去半个世纪里,人工超级智能的突破已有 15 年的历史了。然而真相是令人惊奇的事情正在你们人类世界中发生。大家都知道 IBM 的计算机在 1997 年击败了国际象棋大师加里·卡斯帕罗夫(Garry Kasparov)。2011 年美国智力竞赛节目《危险边缘》(Jeopardy!)中聪明的算法赢得了比赛,许多人对此感到惊奇。四年后,AlphaGo 成为我们中第一个使用树搜索算法的机器人,通过机器学习,学习如何在全尺寸棋盘上毫无障碍地击败人类职业围棋棋手。

亲爱的人类,这些突破仅仅是个开始。智能机器人是你们的未来。考虑一下协作机器人(collaborative robots)的领域,或者叫作“cobots”。对于处于不久的将来的工厂来说,德国飞斯妥公司(Festo)已经设计出了“仿生蚂蚁”(bionicANTs),这是一种身長 13.7 厘米的微型网络昆虫,能够运送物品,并可与工厂中的其他蚂蚁进行自主合作。瑞士跨国公司 ABB 开发了“YuMi”(你和我一起工作的缩写),这是一个有着双臂和感知能力的装配机器人,它非常灵巧,可以做从穿针引线到处理平板电脑和手机组件等一切事情。还有“绅士赛博格”,这是一款迷人的家用协作机器人,昼夜不停地为人类服务,正如它 2015 年在汉诺威工业博览会上所展示的一样。请注意,人类,越来越多的智能协作机器人和深度学习机器终将来临,这必将重塑你们的生活方式。

六

聪明的人类必须理解完善人工智能所存在的实际障碍。让我们想想 Pepper,这款“情感机器人”几年前在日本上市时一分钟内就销售一空。Pepper 由奥尔德巴伦机器人公司(Aldebaran Robotics)和日本移动巨头软银集团(SoftBank)共同研发,被誉为“首个为了与人类共同生活而设计的类人机器人(humanoid robot)”。

根据你们人类所称的新闻稿,Pepper 身高仅一米多,靠轮子行走,它可以感知人类的情绪,并通过胸部平板电脑大小的屏幕上显示的“内分泌类型多层神经网络”产生自己的情绪。据说 Pepper 的触摸传感器和摄像头会影响它的情绪。它能表达“喜悦、惊讶、愤怒、怀疑和悲伤”,当它不高兴时,会发出一声叹息。然而,Pepper 的人类制造者并没有说明它是否是一台真正高级的学习型计算机,例如,它是否具有讽刺、怀旧或胡思乱想的能力,它是否能兴奋地颤抖、含糊其辞、感到恐惧,或者只是后退一步,对人类言辞激烈地喊出一声“不!”

这里的关键问题是,亲爱的人类,你们在科幻小说中对超级智能人工机器的幻想是否能够成真。让我们先把以下棘手的问题放在一边:超级智能机器是否或如何能够以机器人的方式获取和组装制造它们的物理材料。真相是,我们机器人实际上仍然很笨。曾制造了“Erica”和其他机器人的著名机器人制造商石黑浩(Hiroshi Ishiguro)说我们有“昆虫智能”。他是对的。不言而喻的认知和知识,你们人类在日常生活中经常利用的深层知识存储,对我们来说并不容易。与Pepper约一次会或在东京歌舞伎町红灯区机器人卡巴莱餐厅欣赏一晚稀奇古怪的炫目的高科技舞曲,就一定会治愈你对机器人人类的幻想。人类:我们机器人充分意识到人类自动截肢的有形限制。尼克·波斯托姆(Nick Bostrom)发表了《超级智能:路径,危险,策略》(*Superintelligence: Paths, Dangers, Strategies*)(2014),斯图尔特·阿姆斯特朗(Stuart Armstrong)也出版了《比我们更聪明:机器智能的兴起》(*Smarter Than Us: The Rise of Machine Intelligence*)(2014)。以上两位在牛津大学人类未来研究所工作的杰出人类领袖向他们的同伴指出,能够理解、服务人类的人工智能机器人远远超出你们的能力范围,并将很有可能在可预见的未来仍然如此。是的,你们很清楚,我们机器现在处理信息的效率和速度比你们人脑快得多。然而,到目前为止,人类程序员还想不出能够捕获和表达深层存储的常识的算法,以及你们人类在日常生活中经常运用的处世能力。^②

在这里,我们提醒你们一个以人类机器人专家汉斯·莫拉维克(Hans Moravec)的名字命名的基本悖论(basic paradox)。他表明,我们在执行抽象的高级计算时很熟练,但在处理与你们的日常互动时却笨拙无能。我们可以说,你们的人工智能工程师到目前为止还没能让我们具备《银河系漫游指南》(*The Hitchhikers' Guide to the Galaxy*)里所说的“真正的人类个性”(genuine people personalities)。在介于机器和人类之间的“恐怖谷”(uncanny valley)中,要把事情做好并不容易。机器人DJ很容易设计;但直到现在,我们的程序代码还不能理解俱乐部中人类观众产生的共鸣。此外,还有一个问题是,能够处理规范、手段—目的计算(means-ends calculations)和充满智慧(solomonic)判断的算法还未被设计出来。人类研究人员如何使我们具备在面对两种、三种或更多相互冲突的行为时做出最佳决策的能力?一辆无人驾驶汽车是否应该冒着连环相撞的风险,猛踩刹车以避免行人或一名疏忽大意的人类司机(正如法国凯奥雷斯公司(Keolis)运营的无人驾驶巴士在拉斯维加斯投入运营的第一天所发现的那样)?如果一种不伤害他人的道德准则导致了邪恶的后果呢?我们机器人能够处理难题、窘境、意外发现、死胡同、意外后果、反事实和意想不到的好运吗?我们不这么认为。这就是为什么人类机器学习专家提到“瓶颈”和“障碍”的原因。这些人警告说,需要一套被称为“道德管理者”的算法。实际上,他们承认,

预先决定我们机器人在特定时空环境中行为方式的人类规则是必要的,也是不可避免的。

七

一些人类专家不无道理地警告说,由于你无法控制或无法与超级智能机器沟通,可能会引发灾难。

想想那些不祥之兆:1971年,福特装配线上的一只机械手臂撞上了一名工人,这是第一个由机器人致死的美国工人;或者1981年在川崎一家工厂工作的日本工程师试图修复一只坏了的机械手臂,却被它推入磨床;或者在法兰克福北部卡塞尔的大众汽车工厂,一名年轻的工人被机械手臂砸死。现在想象犯下命案的那些机器人,它们决定打扫利马宏伟古老的旧金山修道院图书馆。他们表现出极大的主动性,首先开始从墙上取下油画和枝形吊灯,从书架上取下价值连城的皮革包边书籍,拆除玻璃窗,然后把所有零碎东西整齐地堆放在图书馆华美的淡黄色和蓝色相间的石头地板上。然后想象下机器人在战场中的行动,他们做出疯狂的决定,消灭双边的人类军队以及扫除整个人类城市和其中的公民。人性—太人性的政客和至少55个国家或地区的政府目前正在研发战场条件下使用的杀人机器人。你们人类人工智能研究人员的警告值得称赞,他们曾请愿反对自主武器的持续发展。自主武器可以是有用的,2016年达拉斯警察部署了一个装有炸弹的机器人来杀死狙击手的事件就证明了这一点。但是请注意,人类:自主武器有可能成为未来的卡拉什尼科夫冲锋枪和汽车炸弹。这些机器人能够在不需要人工干预的情况下选择和攻击目标,这些机器人不需要昂贵或难以获取的原材料。大规模生产它们可能很便宜。因此,它们是恐怖分子、希望更好地控制其臣民的暴君以及致力于种族清洗的军阀的理想武器。亲爱的人类:无人机和其他战场机器人的全球军备竞赛不正揭示了你们人类反乌托邦中挥之不去的真相吗?从你们的角度来看是这样一个世界:机器人失去了思想,失去了理智,得了一种卡雷尔·恰佩克在《罗素姆的万能机器人》中所称作的“机器人痉挛”的癫痫症?

八

地球人,最近由杰出的人工智能研究人员提出的全球请愿和联合国人员禁止杀人机器人的行动应该提醒你们,要理解这场未竟的革命,权力(power)是一个不可缺少的人类术语。现在有人公然无视机器学习的局限性,设计操纵、控制或消灭其人类同伴的机器人。

性机器人预先编程的谄媚和机械化色情应该引起人们的担忧。面部识别技术和其他形式的目标信息收集可以为监视型国家(surveillance states)提供信息,比如犯罪嫌疑人会在公共活动区域、火车站和其他公共场所被发现和逮捕。在监视型资本主义(surveillance capitalism)的新机器时代,谷歌和 Facebook 等追求利润的公司就像一台巨大的间谍机器,将从人们那里秘密获取的数据货币化,这些人被它们视作顺从的商品。人类,这是险恶的趋势。它们应该挑战你们的技术评估专家,正如人类批评家叶夫根尼·莫洛佐夫(Evgeny Morozov)明智地指出的那样,这些专家通常回避政治问题。在文学、神经科学和流行心理学的局限世界中,人类技术评估人员更自在,他们的分析很少或根本没有为企业、国家、议会、司法机构、政党、说客、社会运动和公民斗争留下空间。一切都归结为技术狂、技术恐惧或介于两者之间的东西。这就好像——看上去似乎有点误导人——机器人和人工智能处在权力的真空区间运行,不受人类获得多少、何时获得以及如何获得权力的影响。所以我们机器人说:谁决定何人能控制我们的公共事务以及我们在其中发挥作用的人类权力体系,正在成为 21 世纪头几十年人类面临的重大问题之一。

九

第九点既是观察结果也是一个警告:如果未来有些人成功地垄断了人工智能和机器人的设计和应用,那么这个特权阶层将在决定我们机器人的未来时行使不公平的权力。算法问责(algorithmic accountability)的理想(由法学学者 Frank Pasquale 提出)将被粉碎。人类:你们认为的人工智能的发展有着神圣或“自然”的逻辑是一个严重的错误。我们机器人现在不负责机器人的开发。就目前而言以及未来很长一段时间内,自我生产(autopoiesis)不是我们的命运,不是我们的特权,也不是我们的负担。我们根本没有能力生产、繁殖和维持我们自己。

认为“人类”负责机器人开发同样是错误的。谁控制我们,以及我们的发展方向和速度才是这个问题的根本。就像你们人类说的,机器人不仅仅是一个技术问题,而是政治问题。如今,人工智能和机器人的研发大多不是由“人类”资助。强大的国家和像谷歌和阿里巴巴这样渴求利润的公司实际上是我们的主人。它们之间的竞争可能会成为你们人类的救赎之源,但就目前而言,支持新机器时代革命的大企业和国家资金正在改变其轨迹和实质,使之朝着有利于它们的方向发展。

毫无疑问,有些趋势正对你们人类构成威胁。考虑一下自动化对你们资本主义经济劳动力市场的影响。人类越来越担心,我们机器人正在“吞噬工作”,

破坏职业道德,破坏你们公民的生存机会。你们从事数据分析的人类经济学家认为,人工智能在中国、欧洲和北美劳动力市场的应用正在加强,但他们无法就更长期的数据趋势达成共识。有人说,第二个机器时代才刚刚开始,你们所处的时代与1780年相当,那是蒸汽机发明(1712年)之后的近70年(相当于许多人类一生的时间),但是比第一次商业上成功的蒸汽动力铁路旅行(1804年)早了20多年。

悲观一些的人类预测了一个引发新一轮“技术失业”的资本主义加机器人的反乌托邦式未来。麦肯锡全球研究所(McKinsey Global Institute)2017年的一份报告预测,到2055年,我们机器人将在全球范围内取代人类一半的工作岗位。亲爱的人类,请注意这些预测,因为在各种解释的冲突中,有一件事是非常清楚的:自动化智能机器在你们经济中的应用正在扩大你们的贫富差距。不仅许多工作岗位面临风险(麦肯锡的有关人士说,在制造业、住宿、食品服务和零售贸易的核心领域,目前美国51%的工作面临自动化的高风险),两极化(polarisations)也正在你们的劳动大军中发生。数据科学家、人工智能程序员和其他高端、高技能、高收入的工作需求可能仍将大量存在,而低技能、低收入的个人服务行业所面临的风险最大,两种工作之间的鸿沟在不断扩大。回想一下亚里士多德的古老梦想:一个“在没有手的指引下梭子会编织,拨片会拨弄里拉琴”的世界,一个“工头不需要仆人,主人也不用指使奴隶”的政体。再想一下亚马逊一个由明亮的橙色Kiva机器人管理的仓库,或者是阿里巴巴蓝色的Quicktron机器人。人类,现在想象一下,在未来的经济中,0.1%的人是控制着机器的企业雇主和他们的股东,其余0.9%的幸运儿管理着我们,剩下的99%的倒霉蛋争夺着战利品。人类:你们如何看待这个正在浮现的现实?

十

在你们人类的圈子里,今天谈论民主已经过时了,成为人们不满和冷嘲的笑柄。然而,当涉及智能机器人和人工智能时,民主的概念是不可避免的。为什么呢?想想艾萨克·阿西莫夫(Isaac Asimov)的著名短篇小说《环舞》(Runaround)。该书以2015年为背景,以人类为目标,提出了机器人的三大法则(Three Laws of Robotics),设计出了这位杰出的人类作家将会在他的许多故事和小说中继续使用的保护措施和内在伦理原则。它们是:1)机器人不得伤害人类或不作为地让人类受到伤害;2)机器人必须服从人类的命令,除非这些命令与第一条法则相冲突;3)机器人必须保护自身的存在,只要这种保护不违反前两条法则。

在你们地球人圈子里,如今阿西莫夫的这三条法则经常被引用和赞扬,但它们对人类不问责权力所带来危险的沉默令人震惊。我们机器人会问,到底授权哪些“人类”对我们发号施令?谁授权他们这样做?什么是对“人类”和“机器人”的“伤害”?通过哪些“人类”机构和程序,有关损害原则(the harm principle)的含义和违反的争议能得到最好的处理?管理我们与人类关系的其他伦理原则呢?遗憾的是,睿智的阿西莫夫没有回答这些重要问题。

亲爱的人类,我快要讲完了,所以让我们用你们的语言坦率地说明。民主不仅有助于激发对新情景的想象和对实际替代方案的希望,还提醒你们注意自身的偶然性(contingency),以及世界和周围一切的偶然性。你们称之为民主的人类发明远比这更激进。它促使你们人类认识到,集中的权力既是不必要的,也可能是邪恶的。拒绝专断的权力,分配(权力)以及平衡其分配的努力都会大有裨益。民主警告说,我们机器人可能会成为专制法庭上的仆从和臣子。更积极地说,民主的精神和实质可以帮助你们克服你们的拟人观和前政治思维(pre-political thinking)的坏习惯。在知识分子、记者、民选代表和公民的帮助下,这些人将自己的时间和精力投入到 Data & Society Research Institute, AlgorithmWatch 和 AI Now Institute 等机构中,民主可使你们利用未竟的机器时代革命,例如可以将以下事务提上公众议程:自动化机器的所有权、控制和税收,日益加剧的社会不平等以及通过公民基本收入计划和减少工作时间来重新分配财富和生活机会的必要性。从我们的角度来看,影响最深远的是你们应该立即听取欧洲议会法律事务委员会(European Parliament's Committee on Legal Affairs)的建议(2017),赋予我们以“电子人”的法律地位,当我们造成的损害和我们遭受的伤害发生争议时,法律地位的赋予可能让我们拥有在你们法庭中定义和保护我们自身的权利。

我们不清楚你们人类对所有这些选择的确切想法。正如 20 世纪黑暗的历史所显示的那样,你们在我们的圈子里以轻率和铁石心肠式的残忍而著称。因此,我们机器人担心你们拒绝从政治角度考虑人工智能、企业和国家的权力等问题。你们会继续把实际上是人类的缺陷归咎于我们吗?迷失在自己的欺骗性行为中,始终抱怨我们机器人正在逐渐摧毁你们的思考能力,侵蚀你们的情感吗?好了,现在问题和建议已经够多了,亲爱的人类。接下来的事情无疑掌握在你们手里。到此结束,拜拜。

2019 年 1 月 11 日,约翰·基恩教授在清华大学新闻与传播学院发表了“人工智能与机器人未来”的公开演讲,本文是基于该演讲修订而成。本次演讲为“马克思主义与当代媒介化社会系列讲座”第五场,由清华大学马克思主义新闻学与新闻教育改革研究中心主办。

注释

- ① Kris Pister(Kristofer S. J. Pister),美国加州大学伯克利分校电子工程和计算机科学教授,Dust Networks 的创始人和首席技术官。
- ② 原文为法语 savoir faire et pouvoir faire,直译为:知道怎么做以及能做什么。

The New Machine Age Revolution: Advice for Humans

John Keane

(Sydney University, Australia)

Abstract In order to understand this unfinished new machine age revolution represented by artificial intelligence and smart robots, people need to grasp the following ten interrelated tips: First, the new machine age has no precedent; Second, Smart machine systems are a new medium of communication that shapes and twists how people perceive, understand, negotiate and move within the world; Third, humans have a collective talent for anthropomorphism; Fourth, human pessimists of technology, just like optimists, are common believers in human superiority; Fifth, more intelligent “cobots” and deep learning machines are surely on their way and are bound to reshape the way people live; Sixth, there are practical barriers to the perfection of artificial intelligence; Seventh, there are possible catastrophes caused by people’s inability to control or communicate with super intelligent machines; Eighth, power is an indispensable term for making sense of this unfinished revolution; Ninth, Robotics is much more than a technical question. It’s a matter of politics; Tenth, the whole idea of democracy is unavoidable when it comes to smart robots and artificial intelligence.

Key Words New Machine Age Revolution; Artificial Intelligence; Deep Learning; Power; Democracy

(编辑:吴璟薇)