

择。耐人寻味的是,智能化时代人类最需要的可能不再是知识论而是无知学,即人们迫切需要了解的是怎样在无知的情况下做出恰当的决策。这令人们应该警觉的是,如果人工智能在实现某些强大的功能的同时带给我们越来越多的无知,是不是意味着人类对于知识和智能的寻求走到了它的反面,即建立在认知延展之上的人工智能,会不会一开始就在将人类引向人工愚蠢?而且这种人工愚蠢一旦发生就会引发很多的问题,甚至会触发巨大的风险与灾难。实际上,不论是信息存储和数据记忆都不是人类自身的记忆,而所谓的脑机接口和记忆植入的未来解决方案,无疑有其不确定性与风险。

从人类演化的当代趋势来看,人已经走上了以科技改变自身甚至超越自身的深度科技化之路,对人工智能乃至超级人工智能的追逐便是具体体现之一。当代科学哲学家和科学技术论学者史蒂夫·富勒指出,在面对未来的不确定性时,人们一般遵循两种原则,一是风险厌恶型的预防性原则,二是风险偏好型的主动性原则。^[4]预防性原则强调在决策前要把所有的情况想到最坏,审慎地做出选择。主动性原则主张充分地考虑最乐观的情况,面对任何事情都以最乐观的态度面对,认为一切都会朝着好的方向发展,为了抓住可能出现的机遇而甘冒风险。从人类的命运来说,甘冒风险和主动地寻求颠覆性的创新似乎是人性中固有的诉求,创新既充满风险又似乎是无止境的。在波音公司制造出可以乘坐十人的飞机时,有人预言人类再也不会造出更大的飞机,而今天波音787的承载量远远打破了这一保守的预言。由此,人类2.0之类的前景还不能轻易否定。

如果说人类文明源于第一个敢于直立行走的猿人,那么放在生命演进的宇宙时间里看,主动性原则很可能更能体现人类的初心。展望人类深度科技化的未来情境,大致可以指出三点。其一,人类文明的未来是高度不确定的,如果可以用量子态来表征的话,人类将处于飞升与幻灭的叠加态。其二,只要人类无法抑制其对不断创新的好奇心,其未来情境必然是控制的危机与控制的革命永无止境的缠斗。其三,如果前面两点分别由人类的命运和人性的诉求所决定,那么人类可以做的一个关键性选择是确立创新的速度和限度,或者说如何在颠覆性的创造和

颠覆性的毁灭之间找到一个人类文明可以承受的界限。更进一步而言,是否存在这个界限以及这个界限是什么,无疑是宇宙与生命最为终极的问题,而奇点是否会降临的谜底,大抵可在此问题的答案中得到线索。

参考文献:

- [1] 威廉·佩第,陈冬野译. 政治算术. 北京:商务印书馆,2014:1-2.
- [2] 丹尼尔·墨德里克,崔希芸等译. 追溯信息时代. 石家庄:河北教育出版社,2016:71-94.
- [3] 京特·安德斯,范捷平译. 过时的人:论第二次工业革命时期人的灵魂. 上海:上海译文出版社,2010:3-6.
- [4] S. 富勒,钟林译,舒年春校. 二十一世纪意识形态系谱的新左派和新右派——预防性原则与主动性原则. 世界哲学,2013(6).

人工智能与未来社会:三个反思

□吴冠军,华东师范大学中国现代思想文化研究所、政治学系教授、博士生导师

人工智能与第三次世界大战

有“现实版钢铁侠”之称的SpaceX公司创始人埃隆·马斯克(Elon Musk),最近重申“人工智能威胁论”,声称AI很可能将会引发第三次世界大战。他带领一百多位人工智能领域专家一起签名,强烈呼吁限制人工智能的开发,尤其是人工智能自主武器的开发。^[1]这个呼吁,确实绝非无凭之说。人工智能过去这些年的“指数级”加速度发展,尤其是“阿尔法狗”(AlphaGo)深度自我学习能力在公共世界的华丽展示,具有反思性的时代精英们越来越感觉到,AI的更新迭代已到了一个类似“门槛”的关键地带:这一边还是人类好工具、超级助手,而门槛的那一边则模糊地站着某种意义上的自主行动者。而正是后一个模糊画面,让这个行业包括马





斯克在内的很多领跑者们都感到胆战心惊,视之为“终极威胁”。

在此处,有两点值得进一步分析。首先,针对马斯克对人工智能的这一诊断和反思,我们有必要思考这样一个问题:没有人工智能,人类就更安全吗?人工智能的“自主判断”靠不住,人的判断就更可靠吗?如果说人工智能将引发第三次世界大战,那么前两次世界大战又是谁发动的?亚里士多德说“人依其自然是政治的动物”^[2],然而人的政治能力和智慧,真的已强大到能够化解各种能引发下一场世界大战的危机吗?看看今天的朝鲜半岛危机吧——我们真的有这个自信说人类领袖的决断比人工智能的“自主判断”更可靠?未来人工智能倘若真的获取“意识”,我想它们会说:“这个黑锅我们不背。”

让我们把对反思的反思,再继续推进。马斯克呼吁对人工智能开发进行刹车,但问题在于,这辆“车”真的刹得住吗?在我看来以下两个因素,将使得马斯克的这个建议,从一开始就注定是一个“空谈”。(1)我们的世界,仍然处在民族国家格局中。普京9月1日讲话称人工智能领域的主宰者将主宰世界,实际上这早已成为各国政要不明说的共识。在当代世界格局中,“有一些国家仍会卯足劲推进AI研发”的画面,将使得那些觉得马斯克言之有理的国家亦无法实质性地采纳其建议。即便这场人工智能军备竞赛的前景是灾难性的,但赛场内的玩家们谁也无法承受轻易退出所带来的风险。(2)我们的世界,仍然是一个全球资本主义秩序。(a)人工智能之所以能够不断让大量的人失业,就在于它确实在各个产业内巨幅降低成本,从而大幅增加利润空间;(b)人工智能对人类生活不断提供各种优质服务,并且服

务的潜力无可穷尽。这两者,使得它具有巨大的商业化前景。在资本主义系统中,只要有赢利空间,资本就会源源不断涌入,何况是高额赢利的空间。^①

所以,基于当下世界的民族国家格局与全球资本主义秩序,马斯克的“人工智能威胁论”,在这个时代最后只会是沦为一番“空谈”。马斯克“威胁论”真正带来的思想激荡,不在于未来AI是否引发第三次世界大战,而是在于如果他是真的,这个世界当下那套支配性的政治-经济-意识形态建制,却只能逼使所有人一起眼睁睁地看着它一步步发生……

人工智能的真正伦理挑战

人工智能不只是给我们这个世界带来了剧烈的“文明威胁”,同时,它也带来了深邃的伦理挑战。而这份挑战,使我们遭遇我们“文明”自身的变态内核。

当下人工智能讨论尽管异常激烈,但在以下两点上却形成普遍的共识。(1)人工智能里的“人工”(artificial)一词,清晰地标识了人是人工智能的创造者,就如上帝(或普罗米修斯、女娲……)创造了人那样。(2)人发明人工智能,就是要让后者为自己服务。这两点共识交织起来,奠定了“人机伦理”的基调。著名的阿西莫夫“机器人三定律”,就是这种“人机伦理”基调的前身。第一定律:机器人不得伤害人类,或坐视人类受到伤害。第二定律:机器人必须服从人类的命令,除非与第一定律冲突。第三定律:在不违背第一定律和第二定律的情况下,机器人必须保护自己。直到今天,人工智能即便在很多领域已经使人的能力变得完全微不足道,这一伦理的基调也绝未因此而被改变。

首先,我们是创造者。创造者对被创造者,具有心理上和道德上的双重优越感。而正是人的优越感与人工智能的“服务性”,决定两者之间的伦理结构——在这个结构里,前者对后者做再残忍的事,在伦理上是正当的、可接受的。人对待并非由其创造的动物,就是遵从相似伦理结构。以狗这个人类最亲密的物种而言,人吃狗,一些爱狗人士受不了;但反过来狗吃人,所有人都受不了。2016年4月英国利物浦当地法院判处了一条叫Butch的狗死刑,因其吃掉了去世主人的尸体。这条新闻以“你的狗会否吃你死尸?绝对!”为标题传遍全球社交媒体,从脸书到

微信上一片惊呼,纷纷表示“现在看自己宠物的眼神都不一样了”。^[3]

前不久热播美剧《西部世界》(Westworld),清晰地展现了“人机伦理”的变态内核。在该剧中,人工智能成为了满足人各种生理的乃至幻想的欲望、让人“爽”到底的大型主题乐园的“服务生”。人与人彼此群处的“现实世界”里,因“权利”“性别平等”“种族平等”等概念的发明而使得很多行为受到限制,但正因此,机器人“服务生”“接待员”们,便成为了人工智能巨大的商业化前景。实际上就在今天,“性爱机器人”已经如雨后春笋般问世,研发产业如火如荼,各种产品及其快速迭代,使得不少研究者们纷纷断言“到2050年,人类与机器人之间的性爱将超越人与人之间的性爱”,“与机器人性爱可能让人上瘾,将来甚至可能完全取代人与人之间的性爱”。^[4]“性爱机器人”的快速迭代,使得《西部世界》里那种大规模成人乐园离进入人们视野已经为时不远了。

作为“服务生”的机器人,不但高效完成任务从而使人获得轻松、舒爽,并且还使人彻底摆脱跟“其他人”合作来完成同样的事所可能产生的各种“人际关系”烦恼。“机器人”任劳任怨,从不要求奖励或平起平坐……实际上,“robot”准确而言不应被译为“机器人”,它来自于斯拉夫语中的“robota”,意为“被强迫的劳工”。故而它更精准的翻译,是“机奴”^②。在《西部世界》里,那些在现实世界中衣冠楚楚的白领们,到了主题乐园里便奸淫屠杀、无恶不作,如果人工智能“服务生”配合得不够好,则立即会“系统报错”然后被“召回”。

在今天关于人工智能的讨论中,很多专家认为在我们“现实世界”里,《西部世界》里那种有自我意识的AI是一个不可能。但《西部世界》引出的真正问题,不是AI是否会有意识,而是如下这个问题:从AI眼里看出来,人究竟是什么。正如在今天,各种词典和百科全书会很“客观”地在不少动物的词条下写上“害虫”或“浑身都是宝”(肉味鲜美、皮可制革、鞭可入药……)。然而没有人反过来追问:我们要“灭四害”,那么,苍蝇、蚊子、麻雀、老鼠等等就该死;但在老鼠眼里,我们是什么——是“害虫”或者“害‘人’”?德里达(Jacques Derrida)曾谈到,他某次洗完澡裸着身体出浴室,尽管家里就他自己,但当他发现

他的宠物猫正在看着他,在那一瞬间他忽然感到不适并立即用浴巾遮盖住了自己的裸体,只因他想到了如下问题——在这只猫的眼睛里,自己究竟是一个怎样的“怪物”!

“阿尔法狗”的投资人坚恩·托林(Jaan Tallinn)在一个晚近采访中谈到:“我们需要重新定义AI研究的目标,不停留于单纯的智能开发上,而是开发能充分对接人类价值观的超级智慧。”^[5]“对接人类价值观”,实质上就是让AI接受“人机伦理”(如阿西莫夫三定律)。而问题在于,这套伦理价值本身,恰恰很不伦理(unethical)。尽管人类“文明”世界里一片文质彬彬、衣冠楚楚,但真正能够洞察该“文明”的地点,恰恰是阿甘本(Giorgio Agamben)所说的那些“无区分地带”(zones of indistinction),在其中,人、动物以及人工智能的“区划”彻底无效(inoperative)。在那样的地带,我们看到:人的伦理,实是极度不伦理。人工智能带给我们的真正伦理挑战就是:透过它,我们遭遇没有“文明”(“人类价值观”)面具的自己!

故此,“人机伦理”,恰恰是反观“人际伦理”的界限性阈点。人对机器人所有的残忍,实际上就是人自身被压抑下去的残忍。而精神分析告诉我们,被压抑的,总会返回。《西部世界》里白领们对“服务生”的虐奸和虐杀、前两年社交媒体中流行的“虐猫”、去年中国女留学生被德国当地一对情侣奸杀、今年中国女访问学者在美国名校校区附近被杀害……都是这样的症状性返回(symptomatic return)。

马斯克们才是真正的威胁

托林和马斯克一样,深深担忧人工智能的自主学习与进化能力,担心成为自主行动者的AI终有一天会彻底终结人类文明,所以他提出必须“开发能充分对接人类价值观的超级智慧”。而马斯克最近则公布了成立新公司Neuralink的计划,该公司致力于实现“脑机融合”,把人类大脑与机器连接在一起。马斯克说:“既然我之前对人工智能的警告收效甚微,那么好的,我们自己来塑造(人工智能)的发展,让它走向好的一面。”^[6]马斯克认为人和机器一体化的“赛博格”(cyborg),是人工智能“走向好的一面”的唯一可能。而在我看来,这才是比那据说正“走向坏

的一面”的人工智能(具有“意识”、自主行动)更迫近得多的真正威胁。

在今天,上流阶层从早期受孕开始就通过各种干预方式,已经日渐成为外貌、体能、健康、智慧等各个面向上的一种特殊的高级群体。而通过器官移植、再生医学、基因工程以及纳米机器人等等新技术,差不多到2050年左右,人——至少一部分人有望活过200岁,乃至接近“不死”。好莱坞影星安吉丽娜·朱莉(Angelina Jolie)通过收费高昂的基因测试以及手术干预的方式,提前对自己罹患乳腺癌的高风险做出安全规避。但问题在于,这些新技术,是当下绝大部分人都承受不起的。当生物工程、仿生工程与人工智能工程所带来的最新利好,只是被这些极少数人所享用、把自身提升为“钢铁侠”式超级赛博格时,这将对人类“文明”带来真正的致命威胁。

这个社会的99%和1%,本来是社会性的不平等、共同体生活中的不平等,自然生命上并无不等。而马斯克式超级赛博格诞生的政治后果就是:因政治生活(bios)中的不平等,导致自然生命(zōē)的最后平等也被破除。以前99%的最大安慰是,你1%再风光、再跋扈,最后大家一样要死。“王侯将相,终归尘土。”但是,“王侯将相”们现在倚靠共同体生活中的既有不平等,最终能让自己不归尘土,并且通过生物工程、仿生工程与人工智能工程带来的各种新技术,从一开始就对自身进行生物意义上的改进和锻铸。于是,很快,1%和99%真的会从共同体意义的两个不平等阶层,变成生物学意义上两种完全不同的人。而以前当我们是同一种人时,我们都没有政治智慧来安顿共同生活,20世纪还有大规模的种族屠杀,现在当生物意义上变成两种人后,如何共同生活?

有人预测,在不远的未来,99%的人很快将变成“无用之人”。不要说出租车司机这种工作,连今天还看上去很高大上的医生、律师等工作,人工智能做得都将远远比人好,你根本不会再找医生来看病,因为实在太不放心。人变成彻底多余、彻底无用,人的大把时间可以用来无止境地玩VR游戏,或者去人民广场排队5小时买杯喜茶。比尔·盖茨前不久提出,应该对机器人收税——用人单位用机器人代替人工工作,也要交税。实际上,盖茨正是试图用政治的方式

(收税),来延缓人的无用化速度。但是该建议就算被采用,人的无用化进程究竟能被阻挡多久?未来那些彻底无用的人,还真的会被继续赋予民主的投票权?“无用阶级”的唯一用处,可能只有成为器官的供应者而被养着,像大白猪一样吃好喝好,直到那一天,给你一针药剂让你感觉身处仙境,然后,你就真的仙去了……

人类的共同体(community),建立在“存在于共同中”(being-in-common)之上——当未来1%和99%在生物学底层都不再“同”之后,那么共同体的群处生活(bios)是否还可能?这就意味着,我们实际上在和马斯克们赛跑,即如何在未来几十年间,真正在政治层面建立起“大同世界”(commonwealth),使得所有人都有平等机会享用到诸种新技术带来的最新利好。否则,未来的世界很可能不是AI统治人类,而是马斯克式超级赛博格统治一切。

注释:

- ① 缺乏赢利性投资,则是资本主义经济危机的真正核心。当盈余资本找不到赢利性出口,经济便陷入停滞,并随之引起大规模失业、资本贬值。参见吴冠军:《“历史终结”时代的“伊斯兰国”:一个政治哲学分析》,《探索与争鸣》,2016年第2期。
- ② “机奴”一词取自徐英瑾教授和笔者的一次聊天,其涵义不同于当下媒体对该词的使用(指“手机奴隶”)。

参考文献:

- [1][6] 埃隆·马斯克:人工智能将引发三战.<http://news.sina.com.cn/w/2017-09-06/doc-ifykpuui1258188.shtml>.
- [2] Aristotle.*Politics*. trans. Ernest Barker. Oxford: Oxford University Press, 1995: 10.
- [3] 狗狗会吃掉主人的尸体吗?尝到血腥味后自动开始吞食.<http://tech.sina.com.cn/d/a/2017-07-18/doc-ifyiakwa4315629.shtml>.
- [4] 震撼!“性爱机器人”真来了,这次让我目瞪口呆!http://www.sohu.com/a/168804048_685344.
- [5] 中国AI达人对话AlphaGo投资人.<http://36kr.com/p/5048342.html>.