

透明人与空心人:人工智能的发展对人性的改变〔*〕

○ 张宪丽¹, 高奇琦²

(1. 上海政法学院 国际事务与公共管理学院, 上海 201701;

2. 华东政法大学 政治学研究院, 上海 201620)

〔摘要〕善与恶是人性的两个方面,不同外部条件下,人性的善恶表现会有所不同。随着人工智能时代的来临,个人的隐私空间正不断缩小。这使得人们隐藏自己想法的可能性变得越来越低,而透明人将成为人工智能时代的重要特征。在未来,技术会让说谎的人变得越来越没有容身之处,尔虞我诈的计谋人将可能成为历史,而很少或永远不说谎话的空心人将会成为人的主流形态。空心人的流行会使得潜规则或“套路”的作用逐渐减弱甚至消失。在这一背景下,人们会选择更加善良的行为方式。或者说,由于人工智能的发展,人性向善的条件也会越来越成熟。然而,“套路”的终结也产生了另外一个新的问题,即人们可能会对人类存在的意义更加迷茫,这就需要人类做出更加深入的思考。

〔关键词〕人性论;人工智能;透明人;空心人

DOI:10.3969/j.issn.1002-1698.2017.09.007

近年来,人工智能已经成为国内外研究的一个重要主题,关于人工智能的研究成果大量涌现,但其中关于人工智能对人性影响的研究却相对较少。那么,关于人性善恶的讨论在未来的时代条件下会产生哪些变化?或者说,随着人工智

作者简介:张宪丽,上海政法学院国际事务与公共管理学院博士后,讲师;高奇琦,华东政法大学政治学研究院院长,教授,博士生导师。

〔*〕本文系国家社科基金重点项目“全面推进依法治国与国家治理现代化研究”(14AZD133)、霍英东教育基金会第十五届高等院校青年教师基金资助项目“国家治理现代化的评估指标体系研究”(151091)、上海市人才发展资金资助项目“国家参与全球治理(SPIGG)指数的指标与测量”(201473)、上海市教委“曙光计划”项目“执政党与非政府组织间互动机制的国际比较研究”(13SG50)、上海市教委科研创新项目“中国参与全球治理的理论创新研究”(15ZS060)成果。

能和大数据时代的来临,人们会更加走向人性善还是人性恶?当人工智能逐渐获得人类的特性之后,人类存在的意义将会是什么?这些关键问题对于人类发展的重要性是不言而喻的。

一、人性善恶的历史争论

在人类的思想史上,关于人性善恶的争论一直是一个永恒而经典的话题。尤瓦尔·赫拉利(Yuval Harari)在《人类简史》中指出,人类之所以能够把地球上其他大型生物消灭的原因之一,在于人类具有“讲故事”的重要技能。^[1]用更加学术的语言说,这种技能在很大程度上其实就是人类思想的力量。特定时空的思想可以把个体的人更为有效地组织起来,并以群体的方式发挥出强大效应,从而克服自然条件的限制而获得更好的发展。从思想史上看,任何一种思想都是基于人类自身而产生的,其中一个重要的方面,就在于这些思想往往是以人性善恶的假设为基本前提的。

对于“人性恶”的假设,一直是西方思想发展的重要基础。美国著名圣经学家穆尔(G.F.Moore)明确指出:“由于亚当的罪,世人都成了罪人,大家都处在上帝的律法遭到破坏的厄运下。不仅如此,罪不是人的天性,不是来自上帝的创造;但是由于遗传、效法和习惯,罪成了第二天性,而且仿佛成了人所固有的素质,不可能摆脱的东西”。^[2]因此,“原罪说”是基督教人性论的思想来源,而且在很大程度上也构成了西方思想的源头。^[3]由于这种思想的支配性影响,西方几乎所有的社会科学知识都是围绕人性恶的基本假设而展开的,其核心就在于对人性恶的约束与防范。例如,霍布斯在《利维坦》中提出的“自然状态”,即一切人反对一切人的战争,便是基于性恶观念的假设。^[4]在此基础上,社会契约、国家制度和权利观念等才得以发展起来,由此,这种人性恶的假设也构成了西方现代政治学和法学发展的基础。而现代西方经济学也存在“理性人假设”,即假定每个理性的个体都是自私自利地追求个人利益的行为体。^[5]这其实也是人性恶假设的精细版。上述两种观点在很大程度上构成了西方社会科学知识的主轴,很多研究都是在这两点假设的基础上展开的。例如,国际政治理论中的主流学派——现实主义同样也坚持了人性恶的假设,认为国家对权力的追逐是不可避免的。这一点被古典现实主义、结构现实主义以及之后发展出的进攻现实主义、防御现实主义以及新古典现实主义都奉为圭臬。^[6]

然而,人性恶的假设仅仅是人类知识中的一部分,与西方形成鲜明对比的则是中国思想史中关于人性善恶的讨论。中国的传统思想给人性的本质提供了更为丰富的思想资源。其中最为典型的是儒家学说,儒家从最初开始便十分强调人性善的一面,例如,孟子用“恻隐之心”来解释人性之善。^[7]当然,中国思想中也有强调人性恶的理论,其中最为典型的则是荀子的性恶论。不过,尽管荀子强调人的本性是恶的,但是通过后天改造可以去除人性中的恶,树立起道德观念,亦即所谓“化性起伪”。^[8]荀子的学生韩非子将这种理论再次推进,提出了更加赤裸

裸的性恶论。韩非子用一系列论证和寓言故事论证人性恶的一面，并在此基础上提出法家著名的权力至上和专制的思想。^[9]除了这两种典型思想之外，中国历史上的其他一些思想家也对人性进行了较为复杂的分类。例如，董仲舒在《春秋繁露》里便提出了“性三品说”，他将人性分成上、中、下三品，而区分的标准实际上就是善恶的程度。^[10]

有关人性善恶的简单二分法并不能洞察人性的复杂，我们应当认识到，善与恶应当是人性的两个基本方面，更为重要的是，善恶的出现与人所处的外部环境紧密相关。例如，在生存状态极其恶劣的情况下，为了争夺稀缺的资源，人性会偏向恶的一面；而在生存状态大为改善的情况下，人们减轻了对稀缺资源争夺的压力，人性便更加偏向善的一面。观察人类的发展历史就会发现，人类的演进是一个暴力逐步减少的过程，或者说是人性逐渐由恶转善的过程。有学者通过大量的数据统计和分析后，认为过去世界中的血腥与残酷，已经在今天大为减少，今天我们正生活在人类历史上最和平的时代。^[11]

从经验观察来看，人类在近几十年中对暴力的排斥非常明显。例如，在一百年前，战争中战斗人员屠杀平民并不是什么禁忌，但在今天已经受到绝大部分人的抵制。这种变化也引发了关于战争伦理的讨论，例如，施密特在其著作《政治的概念》中就提出了把战斗人员伪装成平民展开战斗的办法。^[12]在早期关于战争的思想著作中，为达战争目的不择手段似乎是一个被认为正确的公理。但是在战争伦理不断发展的今天，暴力的使用被制定了严格的规则。另外一个支撑人性善的重要内容是慈善的兴起。人性良善的社会必然是充满慈善和善行的。^[13]在现代社会中，慈善越来越成为经济社会生活中非常重要的一项内容。一个合格的企业或公民都要花费一定的时间来做慈善，并将其视作现代生活的重要组成部分。而在工业文明之前或早期，当人们还不能解决自身温饱问题的时候，这种现象是不可能普遍存在的。因此，普遍的慈善或公益行为，与后工业社会的出现有着密切关联。罗尔斯的“差别原则”便是在这一背景之下提出的。差别原则的提出作为西方政治思想史上里程碑式的事件，提倡人们在考虑利益时不再以个人的利益为中心，而要以最不利的人的利益为中心。^[14]这一思想的提出，也给现代社会的公平正义提供了新的标准和依据。更为重要的是，这些原则的运用和实施，也可以逐渐引发心理学中的共情效应，进而从道德革命的层面提升人性趋善的程度。

如果说人性向善的趋势需要一定的外部条件，那么由科技革命引发的生产力大发展和生活形态的进步，则为人性展示其善的一面奠定了基础。这一点在人类现代化的过程中表现得尤为明显，而现代化的基础在很大程度上就是科学技术的不断进步。因此，两次工业革命深刻改变了人类的社会形态和物质基础，而以信息技术为代表的新科技革命则在更深刻的层面改变了人类的生存形态。按照这种逻辑，以人工智能为代表的新一轮科技革命，同样也会影响和改变人性发挥作用的外部基础。那么，随着人工智能所带来的这种基础条件的改变，人性

将会产生哪些新的表现形式和变化,则是需要我们认真思考的。

二、大数据时代下的透明人

可以预见到,在未来的大数据时代中,人类几乎没有什么隐私可言。近年来包括各类社交软件在内的一系列移动互联网 APP 的兴起,极大地改变了人类的生活习惯。其中的一个关键变化在于,每个个体都在不断向数据服务提供商提供自己的数据。但是,许多人在这个过程中并没有意识到这种行为对未来产生的重要影响。例如,人们在 Facebook 或微信朋友圈中上传自己的照片,并在与朋友的相互交流中获得愉悦。多数人只是将这个行为视作非常有趣的事情,却没有想到这一行为对其隐私构成的巨大隐患。例如,如果误加了一个怀有恶意动机的陌生人的好友请求,他就会将你的朋友圈信息汇集成关于你和家人的信息图谱,从而利用其中的漏洞而对你的生活造成威胁或伤害。从目前已经发生的类似案件来看,这将会成为一件轻而易举的事情。

目前,人们不仅养成了向服务提供商提交数据的习惯,并且这一趋势正以加速度的方式向前演进。值得注意的是,在你提交自己数据的同时,往往也在提交别人的数据。例如你发布的照片中包含照片本身的信息以及照片的地理数据与时间数据,并且如果你的照片中出现了另外一个人,那么后台的机器就会通过人脸识别获得另一个人的信息。^[15]在这种情况下,尽管你身在现场,但人工智能会比你更了解你所处的环境。更为重要的是,这些信息在很大程度上是可以永久保存的。这种情况很容易让人想起边沁(Jeremy Bentham)发明的全景式监狱和福柯(Michel Foucault)提出的社会大监狱的景象。边沁发明了一种全景敞式监狱(panopticon),以便于管理者密切监督每一个罪犯,因为充分的光线和监督者的注视比黑暗更能有效地捕捉囚禁者的一举一动。^[16]福柯则提出了社会大监狱的说法,意味着在社会之中,监视就像人的毛细血管一样无处不在,而人也在社会规训中做出自己的选择。^[17]这些场景在很大程度上与大数据时代中人的境况非常相近,即人们生活在这种缺乏隐私和充满监视的环境之中。由此,在大数据时代,人的隐私问题就会变成一种迷思。

因此,笔者提出了一种“透明人”的概念。所谓“透明人”,是指通过大数据的比对和信息还原,每一个个体在一系列时间段的所有行为都可以在事后被展现和还原出来,由此每个人的生活状态都将在很大程度上成为透明的。这种情况已经在现实生活中出现了,例如调取监控录像已经成为现代刑侦中的重要手段。当某个争端或犯罪活动出现的时候,四周往往都会存在监控摄像头,而警察可以通过一定程序调出相关录像而获得关键证据。这种图景在未来可能应用得更为广泛,因为在现实生活中,越来越多元的传感器或软件正在不断产生大数据内容,记录下人们生活的方方面面。例如,我们所有的邮件都会存储在服务提供商的数据平台上,我们所有的文字沟通记录,也都会保存在社交媒体软件提供商所提供的信息平台上。如果有正当需要,相关人员可以通过一定程序获取这些资

料。我们已经知道,在现在的公共空间中已经布满了各种传感器。如果说目前我们还有许多信息采集的死角,那么在未来,这种死角会越来越少。因此,这种人的信息的透明化和无隐私化,在物联网时代会更加显著。物联网意味着原本无意识、无感知以及无信息处理的物,通过人们后期的编程和计算而变得更加智能。物联网的发展,也会逐渐进入到人们的私人生活空间中,去获取人们的数据信息。例如智能家居的发展,就是在家庭中部署更多的传感器和信息处理设备,而且这些传感器和设备都需要通过联网来实现其功能。因此,智能家居的使用关键在于其自动化的功能,尤其是依靠服务商后台人工智能的信息处理来牵引。如果智能家居得到普及,那么人类的家庭生活都会暴露在人工智能的监控之下,这会让人隐私空间变得越来越小。

在人们的观念中,家庭是隐私的重要范畴,在家庭中每个个体都是家庭的领主,他们在这片领地中享有完全的隐私。因此,在美国一些州的法律中,不经主人同意私自闯入其住宅,主人甚至可以使用暴力来驱逐,这种暴力被视作合法和正当的。^[18]按照自由主义的理论,人有决定自己命运和重要事项的权利。^[19]然而,智能家居的发展却使得家庭的私人空间的属性变得越来越弱。因为,智能家居公共空间的属性给传感器的布置提供了合法性。而智能家居的管理要交给联网后的控制中心来处理,这就给控制中心或后台技术人员获取个人隐私提供了便利。当然,智能家居只是物联网的一个侧面,未来的一系列相关发展会更具复杂性。

很多历史学家都致力于“还原历史”,但在很大程度上,历史往往是难以被完全还原的。因为历史是由人来记录的,而由于人的主观性的存在,历史的细节和真相很难被完全还原出来。这种历史的不能还原性,被意大利哲学家阿甘本概括为“证言的不可能”。阿甘本用一个极其抽象,同时也少见的例子来论证这一点,他指出,在奥斯维辛集中营,那些看到过最恐怖事件的人是没办法生存下来的,而生存下来的人恰恰是那些没有看到过恐怖场景的人,因此从更为精确的意义上讲,证言是不可能的。^[20]阿甘本的观点在刑事侦查的领域中具有非常重要的启示。当一件凶杀案发生后,警察所搜集的许多证据和证言实际上是碎片化和不完整的,人们很难通过碎片化的信息还原出完整的现场。然而,随着大数据和物联网时代的来临,这种情况可能会发生颠覆性的改变。原因就在于,人为记录的历史将会被由传感器和电子设备所记录的大数据信息所取代,而后的客观性和全面性,会在很大程度上实现具体场景或事件的还原。

当然,这种“透明人”的发展趋势也可能产生一定的副作用,一个异化结果就是,信息的谬误会对人造成长期的不利影响。例如,一次谣言或某一场司法纠纷,就可能把某些人的相关信息暴露在网络之中。然而由于网络的自我繁殖式的传播,这种负面信息就会永久存在于网络空间中。针对这一问题,社会上出现了关于“遗忘权”的公共讨论。大数据时代的一个重要特性是让记忆变成了常态,而遗忘变成了例外。而在人类历史中,遗忘一直是常态而记忆才是例外。人

类克服遗忘的斗争是通过语言、书写和绘画等形式来实现的,人们发明了这些记录人类知识的形式,并不断强化自己的信息提取能力,从而尽可能地克服遗忘。长期以来,能够“青史留名”的总归是少数。但是随着大数据时代的来临,人类社会的记忆变成了常态,而遗忘变成了例外。也就是说,与人相关的所有信息几乎都会被记录下来,没有什么可以被遗漏。然而,人并不是绝对完美的,人的一生中会有无数次的错误,人们获得发展的重要方式就是在错误中学习,并且在很大程度上忘记错误,否则人就会在自我纠结中痛苦终生。从这个意义上讲,对于人类而言,遗忘也是人的一种生存方式,同时也可以被视作一种权利。

但是,大数据时代中的透明人却很难获得这种遗忘和被遗忘的权利。当关于某一个人的不利信息存在于网络之中时,他人的每一次搜索都会加深这种负面印象,导致不良标签经久长存。^[21]尤其是对于刑释人员而言,犯罪的不良记录会长期存在,这会让他入对其产生标签化的刻板印象,由此产生歧视和自然隔阂。^[22]更为重要的是,这种不公平的负面印象会增加这些人改过自新的难度。例如,刑释人员可能会深刻感受到自己将遭受的排斥和歧视,从而增加他们的心理压力和妨碍其社会融入,甚至可能造成其再次犯罪。^[23]正如电影《肖申克的救赎》里面描述的情景,一位名叫布鲁斯的老犯人在几十年中习惯了监狱的生活,却在出狱之后无法适应外面的生活而最终选择自杀。同时,对于那些没有违法犯罪的人而言,由于长久存在的负面信息而导致的他人的歧视和误解,可能会促使其产生犯罪的心理,从而走向违法犯罪的道路。因此,我们需要对这一问题进行深刻的反思和补救。对于那些在互联网上有负面信息的人,当可以证明这些信息是错误的时候,应该可以要求信息服务提供商全面删除这些信息。当然,这种被遗忘的过程在实践中是极其困难的,因为会有太多的人向服务商提出申请,但提供商根本无法有效辨别这些信息的真伪。而且信息服务提供商也只能删除自己平台上的信息,对于已经转载的内容也是无能为力的。这种情况可以清楚表明人们在大数据时代所面临的困境:人们需要通过信息和数据的提供来改善生活的质量,但是每个个体向社会提供的信息越多,他暴露出的瑕疵和错误也就越多,从而使得对其信息的修改和美化就会越来越困难。

三、空心人在物联网时代中的进化

正如前文所提及的,物联网时代是描述未来的一个词汇。而在物联网中最重要的基础条件应当是广泛存在的各类传感器。在物联网时代,传感器的数量将是人类数量的几百上千倍。且不说未来的智能家居和机器人的应用会以几何级的增长速度来增加传感器的数量,单单目前每个人的智能手机上,就已经集成了20多个不同的传感器。我们经常听到和感受到的传感器,就有重力感应器、加速传感器、光线传感器、距离传感器、电子罗盘等等。而目前更高端的智能手机所配备的传感器还有气压传感器、指纹传感器、霍尔感应器、心率传感器、紫外线传感器、血氧传感器等等。传感器作为物联网中的重要设备,既是信息的采集

方，也是解决指令的发出方。人们的日常生活已经和这些传感器密切相关了，而这种关联程度也会随着技术的普及而越来越密切。从目前来看，传感器未来发展的两个重要领域，就是语音识别和人脸识别，而这两个领域都是针对人的。而这些识别技术进一步发展下去，就会与情感计算结合在一起。也就是说，传感器和后台设备不仅会判断出某一个人是谁，还可以通过其表情来分析其心理状态的变化。可以想见，基于海量的大数据，这种情感分析技术会变得更加精准，而这种技术的一个重要应用，就是让测谎仪的使用变得非常普及和便利。^[24]

在科幻或魔幻影片中经常会出现“读心人”，而这在未来可能会变成现实。例如，人的语音信息中包括了非常多的情感因素，人在紧张状态下的语音和自信状态下的语音是完全不同的。而语音识别软件不仅能识别语音信息，还能分析语音背后人的情感变化，这就是测谎仪所希望实现的基本功能。随着此类识别技术的发展，未来的测谎仪可能会变成手机的一种功能。可以想象这样一种场景，未来的你和某人对话时，可以打开软件来实时测试此人所说话的真假。与此相关，人们也可以通过摄像头采集的数据来分析人的微表情，来判断出这个人背后的一些想法。实验证明，人类的大部分行为都受制于规律、模型以及原理法则，而且它们的可重现性和可预测性与自然科学是不相上下的。^[25]因此，未来的技术发展会让说谎的人变得越来越没有市场。当然，这种理想的情况并不是一蹴而就的，在这些技术的应用初期，人们可能会因此而与身边的亲友产生纠纷，甚至断绝关系。但正如囚徒困境中的博弈模型的结果那样，当人们反复多次博弈之后，合作可能会成为最好的结果。在囚徒困境中，囚徒的烦恼在于由于彼此的信息隔阂导致其不能商量，但是技术的发展将会使这一问题得到很好的解决。^[26]如果这种情况成为现实，那么犯罪率在未来的降低也是可以预期的一件事情。阳光是最好的防腐剂，充分的监督可以极大地降低犯罪率，这一点在工业时代就已经得到证明。近期，为了与智能城市的发展相配套，杭州在城市公共空间中安装了更多的摄像头和传感器。现实证明，这对改善杭州的公共秩序和降低犯罪率有很好的辅助效果。在这种技术条件充分发展的条件下，几乎所有的行为过程都可以被还原出来，这就极大地压缩了犯罪分子活动的隐性空间。因此，如今的许多重要案件，都是通过分析摄像头捕捉的信息而破案的。如前所述，福柯从批判的角度来看待监视对人类自由的约束，但从另一个角度来讲，这种监视也极大地限制了人们的不良行为，特别是犯罪行为的发生。

与此相关的另一项重要的科技进展，则是 DNA 检测技术的普及。近年来，利用 DNA 技术破获的重要案件层出不穷，其中一个较为典型的案件就是著名的白银特大杀人案。犯罪嫌疑人在二十多年前连环杀害多名女性，一度造成白银市的社会恐慌，当时警方储存了犯罪嫌疑人的 DNA 信息，但在当时的技术条件下难以排查到嫌疑人。而在多年之后，坚持不懈的警方在 DNA 配对中发现了非常相近的样本，从而找到了犯罪嫌疑人的亲戚，由此顺藤摸瓜，破获了这起长期悬而未决的案件。^[27]同时，前文论及的阿甘本关于“证言的不可能”的论述，

也因为技术的发展而变得可以商榷。长期以来,对于许多刑事案件而言,相关的证据和证言往往是碎片化和不完整的,而且人们也很难通过这些信息还原出完整的现场。人们对福尔摩斯式人物的向往乃至崇拜,实际上源于这种还原的高难度。然而,随着 DNA 检测等技术的发展,还原现场变得更加可能和容易。即便犯罪嫌疑人采取了一系列反侦察的手段,但总是会在现场留下蛛丝马迹,而这些会随着技术的发展而被警方获取。例如,由于生物监测技术的进步,即便犯罪嫌疑人带了手套,但手套与手指之间依然会产生能量的交换,公安部门便可以通过现场遗留的痕迹进行信息还原。更为重要的是,现在很多国家的公安部门正在建立庞大的 DNA 数据库。这不但有利于对犯罪嫌疑人的追踪,还可以提供震慑功能,使得潜在的犯罪心理化于无形。当然,也有人反对政府对公民 DNA 的普遍采集,其理据在于,采集 DNA 就意味着他成为了潜在的犯罪嫌疑人,但这与司法上的无罪推定理论是相悖的。也就是说,对于没有犯罪嫌疑的普通人,不应当像对待犯罪嫌疑人那样,去采集他们的个人信息。

阿甘本非常敏感地结合亲身经历讨论了这个问题。在“9·11 事件”之后,阿甘本受邀赴美国的一所大学讲学,但在机场安检时被要求采集指纹,对此,他以学者身份表达了愤怒。^[28]出于安全的考虑,机场的这种做法在美国是合法的,但是阿甘本则将其视作纳粹虐待犹太人和残杀国民的前置性事件。在阿甘本看来,当年正是在对犹太人和吉普赛人采集指纹之后,希特勒才得以发动种族灭绝的计划,并最后导致了德国乃至世界人民的悲惨命运。从这个角度来看,美国在机场采取指纹的措施与纳粹德国的行为并无区别。^[29]阿甘本的这种观点反映了 DNA 采集等技术应用对人类自由的限制。但从另一个角度来看,机场采集每个人的指纹也反映出现实社会对人性最坏情形的无奈应对,换句话说,就是做最坏的打算,把残酷性置于首位。^[30]同样以安全检查为例,所有乘坐飞机或火车的人都会被要求检查携带的行李物品。这意味着,每个人都被定义为可能携带违禁物品的“嫌疑人”,因为我们无法通过简单的观察就能把那些携带违禁物品的人与正常游客区分开来。由此,把每个人都当作潜在的危险人物的最坏假设就成了当今维系社会安全的最佳制度。^[31]英国著名哲学家大卫·休谟(David Hume)的“无赖假设”可以很好地解释这一制度设计。休谟认为,虽然大多数人都是好的,只有少数人是无赖,但是在制度设计时必须把每个人都当作无赖。^[32]

这一系列论证都指向一个主题,即在大数据和物联网的时代,几乎所有的与我们相关的历史和场景都可以被还原,而人们在这一过程中的错误行为或者犯罪行为都会被其他人或相关机构捕捉到。一开始,这种错误或犯罪的暴露会引起人们的恐慌,但是在经过多次博弈之后,人们会普遍对这些行为越来越谨慎,不再容易说谎或犯罪,即便有人被卷入错误行为,他也会在第一时间去公布自己的心理状态,因为他知道这件事迟早会被发现,只有事先坦白才能降低人们对其错误行为的谴责。假如这种互动模式能够沿着积极的方向长期演进,那么使用阴谋和尔虞我诈的“计谋人”将可能逐渐成为历史,而真诚可爱、很少或永远不说

谎话的“空心人”将会成为人的主流形态。

四、套路的终结与人类独特性的消失

如果空心人成为人的主流形态,那就会导致“套路”的终结。具体来说,在人类的显性规则之外,还有大量的潜规则存在。人们习惯于用“套路”来表示人类社会中的—些潜规则。也就是说,当一个人新进入某个情境中时,对于这些潜规则是不了解的,只有长期生活其中,才会清楚相应的运行规则和规律。在社会学的学术表达中,这种现象也可以被称为默会知识(tacit knowledge)。^[33]

然而,空心人的流行会使得潜规则的意义越来越弱,甚至逐渐消失。在这一背景下,人们会选择更加善良的行为方式,或者说,人性向善的条件也会越来越成熟。弗朗西斯·福山(Francis Fukuyama)在其著作中讨论了信任的问题,他认为信任增加了社会资本,从而创造了繁荣。^[34] 罗伯特·帕特南(Robert Putnam)的《独自打保龄球》与福山的主题是一致的。帕特南认为,美国人的现代生活使得原先的传统共同体在减弱,人们在工业文明早期的小团体中形成的社会资本正在消失,这对美国社会产生了不利的影响。^[35] 因此,帕特南和福山都强调信任和社会资本对于人的重要意义。戴维·布林(David Brin)也在《透明社会》一书中引用被称为“域名女主人”的埃斯特·戴森(Esther Dyson)的话:“真正的挑战不是让所有事保持秘密,而是限制信息的滥用。这也意味着信赖和更多关于信息如何使用的信息……无论他人怎么要求你更具开放性,最合理的方案就是让这种开放性变成互利共赢。信任但是要验证,没有互惠的合作就是支配关系”。^[36] 这几位学者的讨论都指向一个主题,即信任节省了交易成本,当人们之间不信任时,就会设计出复杂的制度来降低风险,而这些都会成为社会成本。但是,如果生活在一个充满信任的社会中,这些制度和交易成本都是可以节省下来的。例如,传统的图书馆需要设计出一套复杂的程序和制度让借书人在规定时间内还书。但随着社会的发展,目前图书馆都简化了相应程序,人们不需要什么手续就可以拿走相应的书。但是这种做法并没有导致书本的减少,反而吸引了人们把自家的书捐给图书馆,这就是相互信任所产生的社会力量。

互联网和物联网的发展逻辑,实际上就建立在这种开放共享的理念之上。一个明显的案例,就是许多黑客所追求的开源软件,正是基于这种理念而得到发展的。例如,2011年,艾伦·施瓦茨(Aaron Swartz)因公开了从期刊储存机构JSTOR得到的大量科学论文而遭到JSTOR的起诉。由于庭审的不利和自身的抑郁症等原因,他于2013年上吊自杀。由此,施瓦茨被视为互联网开放的先行者和牺牲者,并使得大量黑客针对JSTOR发动了各种请愿和攻击,后者也表示歉意并公开了部分数据。^[37] 这一案例也表明,互联网的发展意味着开放和共享理念的实现基础在增强。安德森在《免费》一书中指出,未来的互联网或大数据服务要做到免费,这使得理性世界中的算计变得越来越失效。^[38] 在这种情况下,精于算计和套路的“计谋人”就会变得无用武之地。人们会更加坦然地面对

套路及其影响,用禅宗的话语来说就是:没有套路就是最好的套路。

然而,套路的终结也产生了另外一个新的问题。即人们可能会对人类存在的意义更加迷茫。人类之前骄傲地认为只有人类是有情感的,并且也认为人类思维的复杂性是人和机器的重要区别。也就是说,包括尔虞我诈和算计等在内的“套路”,在很大程度上也是人性的自然反映。但是套路的终结也就标明,人类的独特性也在消失。而就人类的情感而言,人们要借助于人工智能才能来解决社会中极其复杂的问题,而复杂的问题往往也是和情感结合在一起的。因此,人工智能的发展也会使得机器越来越像人,甚至可以让机器人具备一定的感情。在未来,人与人工智能谈恋爱的情况可能出现,而更为可能的情况,则是人借助人工智能来解决心理问题。目前 IBM 研发的 Watson 机器人可以为人们提供非常便捷的法律服务,但是法律问题往往是与情感问题相关联的,因此未来人工智能的法律服务也需要增加情感的选项和内容。由此可知,机器的情感化是人工智能不可避免的发展方向,而与之并行的则是人类的“空心化”。这种情况的出现,就使得我们去思考一个元问题:在放弃了那些被认为人类独有的套路或复杂情感之后,人类还会有什么独特的东西?进而言之,如果说人工智能会从外部条件的角度促使人性趋向善的一面,使得人们变得“透明”和“空心”,那么人类在实现普遍的善之后还会追求什么意义呢?难道人类追求的意义就是指挥人工智能来为人类服务吗?在某种程度上,对于这些元问题的思考是令人烦恼的,但同时又是不得不面对的。对于人类自身而言,人工智能可以改善生存条件,同时也会通过对人性的影响而带来新的问题,这些都是我们需要深入思考的。

五、结 语

综上所述,关于人性善恶的争论是中外思想史的重要主题,但随着现代社会的发展,尤其是近半个世纪以来,人性向善的趋势变得十分明显。在很大程度上,这与经济条件的改善以及道德伦理的发展密切相关。在未来,人工智能、大数据和物联网的发展会使得人性向善的趋势更加明显。在大数据和物联网时代,“透明人”将成为人类的重要生存状态。因为在人工智能深入生活的情况下,人类会变得没有什么绝对的隐私可言。即便是在家庭这样重要的隐私空间中,也会由于智能家居和机器人的应用,而使得隐私的保护变得非常困难。或者说,在新的时代条件下,遗忘变成了例外,而长久的记忆变成了常态。这就使得与人类比较相近的历史可以被完整复述和还原。另外,伴随着人脸识别和语音识别等技术的发展,测谎仪可能会成为非常普及地应用。而 DNA 等生物技术的发展也使得犯罪现场的还原变得非常容易,这会使得犯罪行为大大减少。这些发展都会促使人类越来越少地采取套路、凶残和错误的行为。或者说,人们在全面监视的威慑之下,通过相互之间的多次博弈,往往会采取真诚的策略而放弃相关的“套路”。因此,从人工智能的发展趋势及其对人性的影响来看,人性向善的外部条件会越来越成熟,走向善良和真诚应当是人类不可避免的选择。但是,

这种“套路的终结”也会引发新的悖谬。也就是说,包括尔虞我诈和算计等在内的复杂情感似乎是人与机器的重要区别,或者说是人类独特性的重要表现。但是,如果人们放弃了这些套路,在推动机器逐步走向情感化的同时而让自身走向“空心化”,以此来实现普遍的善,那么届时人们很可能将对人类的存在意义产生新的困惑。这一终极性元问题,将会是人类在推动人工智能发展的时候所必须深入思考的。

注释:

[1][以]尤瓦尔·赫拉利:《人类简史》,林俊宏译,北京:中信出版社,2014年,第24—25页。

[2][美]穆尔:《基督教简史》,郭舜平等译,北京:商务印书馆,1981年,第27页。

[3]丁光训、金鲁贤主编:《基督教大辞典》,上海:上海辞书出版社,2010年,第793页。

[4][英]霍布斯:《利维坦》,黎思复、黎延弼译,北京:商务印书馆,1986年,第94页。

[5][英]亚当·斯密:《国富论》,唐日松等译,北京:华夏出版社,2004年,第13—14页。

[6]于铁军:《进攻性现实主义、防御性现实主义和新古典现实主义》,《世界经济与政治》2000年第5期,第30页。

[7]孟子认为善是每个人凭内心就能判别的,即“恻隐之心,仁之端也;羞恶之心,义之端也;辞让之心,礼之端也;是非之心,智之端也”。参见《孟子》,万丽华、蓝旭译注,北京:中华书局,2013年,第69页。

[8]荀子认为人性是恶的,善是人自觉克服恶本性的结果,即“人之性恶,其善者伪也”。参见《荀子集解:性恶篇第二十三》(卷十七),北京:中华书局,1988年,第434页。

[9]韩非子在面见秦王的时候谈论道:“赏偷,则功臣堕其业;赦罚,则奸臣易为非”。韩非子认为人本性恶,君主当顺应客观形势推行法治。参见《韩非子》,陈秉才译注,北京:中华书局,2015年,第14页。

[10]董仲舒根据善恶把人划分为圣人、斗筭之人和中民三类。他认为圣人在传统儒家思想中是道德至善的化身,历史中曾经存在过具有先验的善,如三皇五帝。斗筭之人作为圣人的对立面则具有先验的恶,已经失去向善的可能性,而中民经过后天的教化依然可以成善,即“圣人之性不可以名性、斗筭之人又不可以名性,名性者,中民之性”。参见苏兴:《春秋繁露义证》,北京:中华书局,1992年,第21页。

[11][美]史蒂芬·平克:《人性中的善良天使:暴力为什么会减少》,安雯译,北京:中信出版社,2015年,第391—392页。

[12]施米特在《政治的概念》一书中对游击队理论进行了详细的阐述,在他的理论中,战斗人员与平民的身份是相互交叉的。[美]卡尔·施米特:《政治的概念》,刘宗昆等译,上海:上海人民出版社,2004年,第288页。

[13]沈涛、陈国庆:《慈善人性与慈善教育》,《教育评论》2013年第3期,第9页。

[14]罗尔斯在正义原则中设立了一种“词典式序列”,其中第二个优先规则即是正义对效率和福利的优先。John Rawls, *A Theory of Justice*, Cambridge: Harvard University Press, pp.302—303.

[15]Face Recognition Homepage, “General Info”, available at: <http://www.face-rec.org/general-info/>.

[16]传统牢狱的三个功能分别是封闭、剥夺光线和隐藏。边沁的全景敞视监狱推翻了传统牢狱的原则,只保留了第一个功能。边沁认为黑暗说到底还是保证被囚禁者的,可见性就是一个捕捉器。Michel Foucault, *Discipline and Punish: The Birth of the Prison*, translated by Alan Sheridan, New York: vintage books, 1991, p.200.

[17]福柯在《规训与惩戒:监狱的诞生》中提出了全景敞视主义这一名词概念,福柯在书中描绘了人因为瘟疫处处被监视的场景,即“在这一空间中,每个人都被镶嵌在一个固定的位置,任何微小的活动都受到监视,任何情况都被记录下来,权力根据一种连续的登记体制统一地运作着,每个人都被不断地探找、检查和分类……所有这一切构成规训机制的一种微缩模式”。Michel Foucault, *Discipline and Pun-*

ish: *The Birth of the Prison*, translated by Alan Sheridan, New York: vintage books, 1991, p.197.

[18]美国《隐私权协议法》于1905年被第一次采用,50个州只有2个州没有在它们的法律中确认隐私权。此外在个别州还有一系列州立法提供了对公民隐私额外的保护。

[19]Robert Nozick, *Anarchy, State and Utopia*, New York: Basic Books, 1974, p.ix.

[20]阿甘本在《奥斯维辛的残余》一书中论述了这种观点。Giorgio Agamben, *Remnant of Auschwitz: The Witness and the Archive*, trans. by Daniel Heller-Roazen, New York: Zoon Books, 1999, p.34.

[21]Link and Phelan, "Conceptualizing Stigma", *Annual Review of Sociology*, 2001, Vol.27, pp.363-385.

[22]Henderson, "New Frontiers in Fair Lending: Confronting Lending Discrimination against Ex-Offenders". *New York University Law Review*, Vol.80, No.4, 2011, pp.1237-1271.

[23]Olphen and Eliason, eds., "Nowhere to Go: How Stigma Limits the Options of Drug Users after Release from Jail", *Substance Abuse Treatment Prevention and Policy*, Vol.4, No.2, 2009, pp.1-10.

[24]Elena Rusconi and Timothy Mitcheener-Nissen, "prospects of functional magnetic resonance imaging as lie detector", *Human Neuroscience*, Vol.7, No.53, 2003, p.1.

[25][美]艾伯特·拉斯洛·巴拉巴西:《爆发:大数据时代预见未来的新思维》,马慧译,北京:中国人民大学出版社,2014年,第12-13页。

[26]郑也夫:《走出囚徒困境》,北京:光明日报出版社,1995年,第201页。

[27]李兴:《白银杀人案与DNA检测》,《百科知识》2016年第21期,第23-24页。

[28]阿甘本认为:“我们不应该被那些安全的理由所困扰,它们总是被它们的正当性证明推到我们面前。经验曾教会我们,这些实践刚开始仅仅适用于外国人,之后会逐渐运用于每个人。”Giorgio Agamben, "Bodies without Words: Against the Biopolitical Tatoo", *German Law Journal*, Vol.5, No.2, 2004, p.169.

[29]阿甘本在文章中说到:“我想再提醒诸位,奥斯维辛的纹身看起来无疑是对运进集中营的个人进行登记、注册和管理的最正常也是经济的方式,今天强加在我们身上的生命政治纹身很可能在明天成为一种工具。”Giorgio Agamben, "Bodies without Words: Against the Biopolitical Tatoo", *German Law Journal*, Vol.5, No.2, 2004, p.169.

[30]Judith Shklar, *Ordinary vices*, Belknap Press of Harvard University Press, 1984.

[31]何包钢:《民主理论:困境和出路》,北京:法律出版社,2008年,第10页。

[32]大卫·休谟认为:“在设计任何政府体制和确定体制中的若干制约、监控机制时,必须把每个成员都设想为无赖之徒,并设想他的一切作为都是为了谋求私利,别无其他目标”。这是一种人性恶的基本假设。同样,在机场或车站设置安检也是基于一种人性恶的基本预设。参见大卫·休谟:《休谟政治论文选文》,张若衡译,北京:商务印书馆,1993年,第27页。

[33]默会知识是由波兰尼提出,主要指相对于显性知识而言的。它是一种只可意会不可言传的知识,是一种经常使用却又不能通过语言文字符号予以清晰表达或直接传递的知识。Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago: University of Chicago press, 1958, pp.95-97.

[34][美]弗朗西斯·福山:《信任:社会美德与创造经济繁荣》,彭志华译,海口:海南出版社,2001年,第11页。

[35]Robert Putnam, *Bowling Alone: The Collapse and Revival of American Community*, Washington: The Brookings Institution Press, 1999, p.149.

[36][美]玛蒂娜·罗斯布拉特:《虚拟人:人类新物种》,郭雪译,杭州:浙江人民出版社,2016年,第192页。

[37][以]尤瓦尔·赫拉利:《未来简史》,林俊宏译,北京:中信出版社,2017年,第346-347页。

[38][美]克里斯·安德森:《免费》,蒋旭峰、冯斌等译,北京:中信出版社,2013年,第85页。

[责任编辑:汪家耀]