

从艾伦·图灵到伊恩·麦克尤恩： 《像我这样的机器》中的人工智能、谎言与伦理

尚必武

【摘要】本文以《像我这样的机器》为研究对象,审视英国当代知名小说家伊恩·麦克尤恩与人工智能之父艾伦·图灵关于人工智能与机器人的跨时空对话。文章在分析机器人亚当介入人类生活后所引发的冲突与矛盾的基础上,重点讨论麦克尤恩对图灵测试的文学重写,即他巧妙地将“机器能否思考?”这一图灵式问题改为“机器能否说谎?”的麦克尤恩式问题。文章认为,亚当与查理在米兰达说谎一事上的不同立场,折射出机器人和人类之于真相与谎言的不同诉求。与阿西莫夫的机器人学三大法则相左,麦克尤恩暗中认同图灵对查理毁灭亚当这一行为的批判,由此引入了人类是否要为机器人的生命担负责任这一值得深思的命题。

【关键词】伊恩·麦克尤恩;图灵测试;人工智能;谎言;伦理

【作者简介】尚必武,上海交通大学外国语学院教授,博士生导师,教育部青年长江学者,欧洲科学院外籍院士,研究方向:叙事学、文学伦理学、英美文学(上海 200240)。

【原文出处】《外语教学》(西安),2020.3.103~107

【基金项目】本文系国家社会科学基金重大项目“当代西方叙事学前沿理论的翻译与研究”(项目编号:17ZDA281)的阶段性研究成果。

1. 引言

2018年10月,伊恩·麦克尤恩(Ian McEwan)作为“21大学生国际文学盛典”的年度致敬人物首次访问中国,并在颁奖典礼上发表了题为“如果有一天‘人造人’写出了小说”的演讲。在演讲中,麦克尤恩聚焦数字革命,大谈“人工智能”(artificial intelligence)和“人造人”(artificial human)。他说:“许多个世纪以来,在许多种不同的文化中,人们的脑海里一直萦绕着一个梦,那个梦就是创造出一个人造版的我们。就像基督教的上帝用黏土造出第一个人那样,我们自己或许有朝一日也能成为上帝,造出我们自己的第一个人造人”(麦克尤恩2019:65)。麦克尤恩在演讲中所重点讨论的人造人正是其新作《像我这样的机器》(*Machines Like Me*, 2019)的一个重要命题。小说背景放置于20世纪80年代,第一批机器人亚当、

夏娃被投放市场。主人公查理购买了与人工智能之父艾伦·图灵(Alan Turing)同款的机器人亚当。作品以人物叙述的方式,讲述了亚当如何介入查理及其女友米兰达的生活。亚当在同米兰达发生性关系后,声称自己无可救药地爱上了她。最终,他由于揭发米兰达曾经因为替其好友玛利亚姆复仇而在法庭上说谎与诬告的违法行为,为查理和米兰达所不容,被查理用一柄铁锤结束了生命。

马歇尔·塞洛克斯(Marcel Theroux)评论说:“尽管这部小说涉及诸多主题,如意识,偶然性在历史中的作用,人工智能,被忽略的文艺复兴散文家威廉·康沃利斯爵士,俳句的形式要求,计算机科学中的P与NP问题等,但它的真正主题其实是道德选择”(Theroux 2019)。塞洛克斯口中的“道德选择”(moral choice)被斯图亚特·米勒(Stuart Miller)具化为小说人

物在道德决定上的不同立场,即“米兰达和查理最初对于亚当是一个机器还是一个感知存在,产生了分歧。后来,小说的一个关键点是米兰达和查理在凝视亚当的时候,存在一个道德分水岭。”对此,麦克尤恩的回应是,“小说末尾,你认为亚当是一个冷血无情的机器还是一个感知的存在?这是我们所遇到的问题,它为我们探讨道德问题开辟了一个新领域”(Miller 2019)。无论是塞洛克斯还是米勒都敏锐地捕捉到小说的人工智能与机器人题材所触及的道德问题。与之类似,尚必武从文学伦理学视角详细分析了亚当与米兰达“偷吃禁果”和亚当之死两个叙事片段,认为小说“揭示了科学选择和伦理选择之间的冲突,隐喻了具有伦理意识的脑文本在人类生活的无可替代性。即便身处人工智能时代,我们也不能指望机器人来帮助我们解决所有的问题,尤其是伦理道德领域的问题”(尚必武 2019: 73)。上述论者紧扣作品的道德主题,切中肯綮,富有洞见。遗憾的是,他们都忽略了图灵在小说中作为虚构人物的存在,尤其是图灵之于推动作品叙事进程的作用。实际上,从查理购买机器人亚当,到后来亚当与米兰达发生性关系并宣称自己对米兰达产生爱情,再到最后亚当被毁,这些贯穿小说始终的核心事件都在图灵那里得到了评论与解释。一方面,查理关于机器人的困惑得到图灵的耐心解答;另一方面,查理毁灭机器人的行为也遭到图灵的严厉指责。本文试图围绕查理与图灵的交往、麦克尤恩对图灵测试的文学改写、以及机器人的毁灭与人类的伦理责任等话题,考察麦克尤恩与图灵之间关于人工智能和机器人的跨时空对话。

2.“如谜的解谜者”:从现实走向虚构的图灵

2013年12月24日,英国女王伊丽莎白二世较为罕见地颁发了皇家赦免(Royal Pardon),而赦免对象正是于1954年6月在家中吞噬含有氰化物的苹果而中毒身亡、年仅41岁的世纪天才艾伦·图灵。女王的皇家赦免一方面固然是以官方的形式向图灵在晚年因其同性倾向而遭遇的司法不公表示歉意,另一方面也是在向这位二战期间破解德国军事密码、立下

赫赫战功的伟大科学家致以敬意。今天,普通民众对图灵的认识主要归功于安德鲁·霍奇斯(Andrew Hodges)那部被广为流传的传记《艾伦·图灵传:如谜的解谜者》(*Alan Turing: The Enigma*)。该书自1983年问世后,被多次再版印刷,后来直接成为电影“模仿游戏”(*The Imitation Game*, 2014)的脚本。为纪念图灵诞辰100周年,霍奇斯还联手数理逻辑学家巴里·库珀(S. Barry Cooper)策划一部向图灵遗产致敬的文集,试图邀请全球顶尖科学家来简述图灵的科研工作,分享他的思想精髓。经过将近10年的准备与努力,这部题为《永恒的图灵:20位科学家对图灵思想的解剖与超越》(*The Once and Future Turing: Computing the World*, 2016)的文集最终由剑桥大学出版社出版。在该文集中,来自麻省理工学院、剑桥大学、斯坦福大学等全球知名学府的20位重量级科学家从算法逻辑、密码学、形态发生学、素数等多个方面阐述图灵的思想与贡献,并预测相关领域的未来发展。对于当下人们对图灵与日俱增的热情,两位编者颇感欣慰。他们说:“今天的情况截然不同,这么多科学家对本书的热情回应反映了人们对于图灵的极大兴趣,包括他的为人以及所做过的事,与图灵相关的一切都成了人们的关注点。真难想象图灵会怎样看待自己的‘复兴’或者在科学和历史中的地位”(库珀、霍奇斯 2018: 4-5)。

如果说霍奇斯、库珀等人的工作只是代表科学家在向图灵致敬的话,那么我们不禁要问,作家如何以自己的方式向图灵致敬?麦克尤恩的回答是“假如图灵还活在20世纪80年代……”。在接受访谈时,麦克尤恩直言:“为了这部小说的目的,我重新创造了过去。我们在科学史上的位置完全是随意和视情况而定的。可能早,可能迟,也可能在别的地方。所以,让我们把科学在时间上向前推,我们让图灵没有自杀”(Aron 2019: 42)。由此,被复活后的图灵就这样进入了麦克尤恩所编织的故事世界,成为小说《像我这样的机器》中的一个重要人物。在作品中,主人公查理与图灵一共有过三次会面,而会面的主要缘由和话题都是机器人。

查理和图灵的第一次见面是在某个餐厅。当天是查理的生日,米兰达和查理准备在餐厅庆贺一番,而彼时的查理却心神不宁。一个原因是他此前发现了机器人亚当与米兰达两人之间的性关系。在质问亚当的时候,查理被亚当打得手腕骨折,而且亚当还辩解说是自己无可救药地爱上了米兰达。另一个原因是当天查理在餐厅遇到了自己的偶像图灵。前一个原因让查理心烦意乱,备受挫败;后一个原因让查理变得兴奋异常,心不在焉。在查理看来,唯有图灵才有可能解开自己关于机器人亚当萌发爱情的困惑。在现实世界中,图灵是一位知名的解密大师;在麦克尤恩创作的小说世界中,他被寄希望于解开查理心中的秘密。

与初次见面时的拘谨陌生相比,两人在第二次会面时,没有经过太多的寒暄,就直奔主题:“那么,查理,请你告诉我关于你的亚当的所有一切”(McEwan 2019: 173)。在图灵的要求下,查理讲述了除了米兰达所卷入的案件之外的所有事情,包括亚当第一次苏醒萌动到第一次不服从,亚当与米兰达的性关系,亚当对男孩马克的嫉妒,亚当如何打得自己手腕骨折,亚当宣布对米兰达的爱情,亚当关于俳句的理论,以及亚当关闭死亡开关等。听完查理的叙述后,图灵的反应是,“在听到的所有关于机器人亚当与夏娃的情况中,你的机器人是唯一一个宣称自己坠入爱河的。这个值得注意。他也是唯一一个拿暴力开玩笑的机器人”(McEwan 2019: 176)。在图灵看来,亚当是一个有感知、学习能力强、高度智能化的机器人,这是他宣称爱情、嫉妒马克的主要原因。不过,正是在和图灵的这次会面中,查理得知尽管第一批机器人大多都成功关闭了他们的死亡开关,但不可思议的是,也有一部分亚当和夏娃开启自毁程序,表现出自杀的倾向。对此,图灵做出了两点解释:第一,机器人的“这种智能并不完美”(This intelligence is not perfect)(McEwan 2019: 179);第二,智能化的机器人被带入了一个并不完美、充满矛盾与问题的世界。图灵说,“我们创造了智能和具有自我意识的机器人,然后将它推到我们这个不完美的世界。机器

人的心理通常是根据理性的线索,对他人有好感的标准设计的,这样的心理很快就会遭遇矛盾的飓风”(McEwan 2019: 180)。如果图灵的解释是对的,那么便产生了另一个更加让人疑惑的问题:机器人如何凭借其并不完美的智能在这个不完美的世界与人类相处?或许正是在这种意义上,麦克尤恩以文学的方式重写了“图灵测试”。在与查理第三次见面中,图灵重点回应了这一测试结果。对此,本文在第四部分会有详细阐述,此处不赘。

3. 被改写的“模仿游戏”:图灵测试的文学外衣

在接受访谈时,麦克尤恩直截了当地说,“我喜欢这样的—个想法,把图灵测试推向极致”(Aron 2019: 42)。J. M. E. 海兰(J. M. E. Hylan)在“被遗忘的图灵”一文中写道:“图灵大概不会认为自己在百年纪念日之时会被当作国家战争英雄,但是他一定知道自己留下了一笔技术遗产。图灵机和图灵测试是这一遗产的著名部分”(海兰 2018: 37)。这个被海兰所推崇、被麦克尤恩所试图推向极致的图灵测试又名“模仿游戏”(imitation game)。1950年,图灵在哲学期刊《心灵》(*Mind*)杂志上发表了一篇题为“计算机与智能”(Computing machinery and intelligence)的论文,首次提出了图灵测试的理论。图灵将论文第一部分命名为“模仿游戏”,大胆提出“机器能否思考?”(Can machines think?)的问题。论文伊始,图灵写道:

我提议考虑这样一个问题,“机器能否思考?”回答这个问题,需要开始界定“机器”与“思考”的意思。这些定义或许要被框起来,这样可以反映这些词语被正常使用的可能范围,但是这种态度是危险的。如果通过考察“机器”和“思考”的一般用法所得出的意思,则难以脱离通过数据统计和民意调查来寻找“机器能否思考?”这个问题的意义和答案的窠臼。但这种做法是荒诞的。我不试图提供一种定义,而是用另一个问题来替换这个问题,即一个紧密相关、用词表达相对清晰准确的问题。

这个问题的新形式可以用我们称为“模仿游戏”(imitation game)的术语表示。这是一个可以由三个

人来玩的游戏,一个男性(A),一个女性(B)以及一个既可以是男性也可以是女性的提问者(C)。游戏的主题就是提问者需要决定其他两人谁是男性、谁是女性,其判断主要是根据X、Y的标签,游戏的最后,他可以指出“X是A,Y是B”或者“X是B,或Y是A”。(Turing 1950:433)

在图灵测试中,上文中的男性和女性被换成了机器和人。在测试者与被测试者(一个人和一台机器)被隔开的情况下,通过一些装置(如键盘)向被测试者随意提问。如果机器让参与者做出超过30%的误判,那么这台机器就通过了测试,被认为具有人类智能。图灵认为,“我相信在50年之内,计算机的存储容量会达到109,并且能够在模仿游戏中取胜。普通水平的猜测者,在经过5分钟的提问之后,猜对的机会不会高于70%。我相信到那个时候,‘机器能思考吗?’这个问题就会自然地失去意义,根本不值得讨论”(霍奇斯2017:541)。

图灵测试一直是计算机科学和人工智能研究界热议的一个话题。2014年6月7日在英国皇家学会举行的“2014图灵测试”大会上,英国雷丁大学发布新闻稿,宣称弗拉基米尔·维西罗夫(Vladimir Veselov)创立的人工智能软件尤金·古斯特曼(Eugene Goostman)成功通过了图灵测试。那么在小说《像我这样的机器》中,麦克尤恩又是如何回应图灵测试,并以此与图灵展开对话呢?对此,麦克尤恩的做法是改写图灵的模仿游戏,提出了“机器能否说谎?”(Can machines lie?)这个新问题。某天早晨,亚当在充电完毕后,突然对查理说:

“根据我在过去几秒的研究和分析,你应该小心,不能完全信任她。”

“什么?”

“根据我的……”

“你把话说清楚”

我愤怒地盯着亚当苍白的脸。他平静地用悲伤的声音说“她可能是个说谎者,是一个系统的、恶毒的说谎者。”(McEwan 2019:30)

随着故事的展开,米兰达说谎及其谎言背后的

缘由被一一披露。事情真相是,一个名叫科林奇的男子并没有强暴米兰达,而是强暴了她的好朋友、来自巴基斯坦的女孩玛利亚姆。因为担心家庭生活和家族名誉受到影响,遭到强暴后的玛利亚姆没有选择报警,最后割腕自杀了。为了报复科林奇,米兰达故意让科林奇和自己一起参加派对。在发生关系后,米兰达指控科林奇强暴自己,最终导致他获刑入狱,成功实施了为好朋友玛利亚姆复仇的计划。

笔者之前指出,“米兰达对科林奇采取的报复行为,涉及法律和伦理两个维度。从法律层面上来说,米兰达确实做了伪证和诬告,违背了法律,但是从道德层面上来说,她无疑伸张了正义,做了伦理正确的选择。换言之,米兰达的行为,违背法律但合乎伦理”(尚必武2019:69)。作为机器人,亚当难以理解和应对米兰达的说谎行为及其同时涉及的法律与伦理这一异常复杂的情况。他凭借自己通过机器学习所获得的法律知识,只知如何解决法律问题,而不知如何理解和应对道德问题。在这种情况下,亚当将米兰达和科林奇之间的谈话录音作为证据提交给了法庭。

亚当从法律层面指责米兰达的复仇行为,无视米兰达说谎行为的伦理维度,罔顾她即将要领养孤儿马克这一事实,声称“原则是原则”(The principle stands)(McEwan 2019:276),坚持认为她必须接受法律的制裁。尽管查理一再提醒亚当需要考虑玛利亚姆被科林奇强暴并导致她抑郁自杀的这一实际情况,即米兰达的说谎在本质上属于伸张正义的行为。查理与亚当之间关于真相与谎言的对话进一步强化了麦克尤恩对图灵测试的重写。查理有意强调“真相不是一切”(truth isn't always everything),而亚当则振振有词:“真相就是一切”(truth is everything)(McEwan 2019:277)。在真相与谎言这个问题上,亚当显然没有通过麦克尤恩所重写的“图灵测试”,因为机器人既不会说谎,也无法理解人类的谎言。在同查理的第三次见面中,图灵解释说:“我们还不知道如何教机器人去说谎”(We don't yet know how to teach machines to lie)(McEwan 2019:303)。换言之,

说谎已经超越了机器人的生存法则。由是说来,尽管亚当通过了图灵测试,成为一个会思考的智能机器人,但他没有通过麦克尤恩所改写的测试,即他不是一个人会说谎和理解谎言的机器人。

在与人类的沟通交流中,亚当难以区别真相与谎言之间的界限,一味地以机器人的理性,对抗、否认人类的情感与非理性,激化了他与查理和米兰达之间的矛盾。亚当之于谎言的立场和态度,表现出他与人类之间的区别。或许,他所暴露出的问题也可以在图灵缺陷那里得到合理的解释。霍奇斯说:“图灵缺陷的本质就是,把他自然科学的分析方法,应用到了人类社会。所谓的‘客观事实’,也许在研究质数时很有用处,但却不能简单地用来研究人类”(霍奇斯2017:666)。

4. 机器人的毁灭与人类的道德责任:从阿西莫夫、图灵到麦克尤恩

在《谁为机器人的行为负责?》一书中,乌戈·帕加罗参照机器人科幻小说之父艾萨克·阿西莫夫的观点,认为我们需要通过法律的形式来界定机器与人类之间的关系。帕加罗(2018:18)说:“在法律与文学之间,阿西莫夫故事中的信息看起来很清晰:既然机器人要留在这里,法律要做的应该是明智地管理我们彼此的关系。”众所周知,阿西莫夫在短篇小说《转圈圈》中提出了机器人学三大法则:1)机器人不得伤害人类,或因不作为而使人类受到伤害;2)除非违背第一法则,机器人必须服从人类的命令;3)在不为违背第一及第二法则的情况下,机器人必须保护自己(阿西莫夫2014:201)。在小说《像我这样的机器》中,亚当至少有两次侵害或威胁人类的行为。第一次是在他和米兰达偷吃禁果、发生关系后,将前来兴师问罪的查理打致手腕骨折。亚当甚至还警告查理不准再碰他的死亡开关,否则会直接卸掉查理的胳膊。第二次是亚当在掌握米兰达说谎的证据后,将它交给法庭,导致米兰达被捕入狱。根据机器人学的三大法则,亚当显然违背了机器人伦理,对人类构成了威胁。

阿西莫夫关于机器人学的三大法则在很大程度上

上成为讨论机器人伦理和人工智能伦理的一个固定程式。在《像我这样的机器》中,麦克尤恩反其道而行之,提出了人是否要为伤害和毁灭机器人承担责任这个问题。由于亚当坚持要举报米兰达说谎的违法行为,查理最终做出了毁掉亚当的决定。问题在于,在人工智能时代,查理是否要为毁灭机器人的行为承担法律责任和伦理责任呢?这也是查理和图灵在第三次见面时所讨论的核心内容。

临终前,亚当请求查理和米兰达不要将自己的躯体交还机器人公司回收,而是送给图灵保管和研究。在图灵实验室,得知亚当死亡真相后的图灵严厉批评查理:

我希望有一天,你用锤子对亚当所做的一切会构成一种严重的犯罪。是因为你购买了他?你有这样权利与资格?(McEwan 2019:303)

你绝不是一个被宠坏了的孩子,在打碎自己的玩具。你不是单纯地在否定法律规则的一个重要观点。你在试图摧毁一个生命。他有知觉,有自我。他如何产出神经元、微处理器和DNA网络,并不重要。你认为只有我们才有特殊的礼物吗?问任何一个狗的主人。弗兰德先生,这是一个好的大脑,我怀疑它比你我的大脑都要好。这是一个有意识的存在,你却努力去清除它。我很鄙视你的这种行为。(McEwan 2019:303-304)

图灵将亚当看成一个人格化的机器人,认为亚当拥有自我和意识。换言之,在图灵眼里,亚当是一个有生命的个体。他可以意识到自己的存在,尤其就智力而言,亚当的大脑超过了图灵、超过了查理。在这种意义上,查理毁灭亚当与一个被宠坏的孩子肆意破坏自己的玩具,有着本质的区别。值得注意的是,图灵在这里特意使用了“意识”(consciousness)这个词汇。人工智能领域的一个共识是,机器人一旦有了意识,就不再纯粹是机器。斯图亚特·拉塞尔(Stuart J. Russel)和彼得·诺维格(Peter Norvig)直接指出:“如果机器人有了意识,那么把它们当作纯粹的‘机器’来对待(例如将它们拆开)可能是不道德的”(Russel & Novig 2013:867)。由此不难理解图灵为何

不仅在道德上批判查理毁灭亚当的行为,而且还认为他的行为是一种犯罪。面对图灵的严厉指责与批评,查理似乎受到了触动,开始为自己毁灭亚当而感到愧疚。作品将查理的愧疚感展现得淋漓尽致的一个细节是他在趁着图灵因故短暂离开的间隙,偷偷逃离了图灵实验室。

在离开图灵实验室之前,查理对亚当的躯体做了一场耐人寻味的告别:

我站在亚当旁边,把手放在他上衣的翻领上,放在已经停止跳动的心脏上。脑子出现一个不着边际的想法,布料很好。我斜趴在桌子上,低头向下看着他那双无光、浑浊的绿眼睛。我没有特别什么特别的意图。有时候,身体比大脑要早一步知道该怎么做。尽管他或许给马克带来了伤害,我假定自己认为原谅他是正确的。希望他或者他记忆的继承者可以原谅我和米兰达的可怕行为。犹豫了几秒钟之后,我把脸贴在他的脸上,吻了他柔软、比人类更像人类的嘴唇。(McEwan 2019:306)

在上述关于查理向亚当遗体告别的片段中,叙述者有意强调了“原谅”这一词汇。一方面,他认为自己已经原谅了亚当,尽管其行为会给马克带来伤害,但实际上米兰达才是最直接的受害者。另一方面,查理希望亚当也可以像他们原谅他一样,原谅自己,而且还承认自己的行为“可怕”(terrible)。这是叙述者第一次使用“可怕”一词来形容自己用锤子摧毁亚当的行为,说明他在道德立场上承认自己有错,尽管并没有法律条文可以判定查理有罪。其实,彼时查理的心态是矛盾的:在言辞上,他不愿意承认自己有意向亚当的躯体认错,但他内心又渴望能和亚当达成和解、原谅彼此。查理的这种矛盾心态主要归因于他把亚当既看作是人、又看作是机器的模糊立场。譬如,在俯身亲吻亚当的时候,查理突然意识到亚当终究不是人类,其原因带有明显的悖论性质:亚当比人类更像人类的嘴唇背叛了他的机器人身份。

该如何解释图灵对查理的严厉指责以及查理之于机器人亚当的矛盾心态呢?或许,答案就在瑞恩·

卡洛等人所说的“机器人人格化的陷阱”。在《人工智能与法律的对话》一书中,卡洛等人认为再智能化的机器人也是机器,人类需要谨防落入机器人人格化的陷阱,指出:“我们必须避免人形机器人陷阱。机器人仅是机器,即使那种复杂的机器人也不例外。在可以预知的未来,它们也不可能超越机器的范畴,我们应依此做出相应的立法。陷入(机器人)人格化的陷阱,将导致我们陷入自相矛盾的处境。”(卡洛,等2018:20)循着卡洛等人的逻辑与思路,我们大致可以判断在小说中,图灵基本上完全落入了机器人人格化的陷阱,而受到图灵影响,叙述者查理也开始表现出将机器人亚当人格化的倾向,不过他的这一态度转变得并不彻底。

史蒂芬·卢奇和丹尼·科佩克(2018:452-453)预言:“机器人不再只是人工智能的早期历史中一个未来主义的话题,它们是生活的现实,并且逐渐成为人们日常生活的一部分。”既然与机器人共处是未来社会生活的一个常态,那么人类该如何面对这一具体现实?对此,小说家又做出怎样的文学回应?通过聚焦图灵与查理围绕机器人亚当之死这一事件的对话,麦克尤恩试图将人类与机器人之间关系的讨论引入道德领域。麦克尤恩认为,这是小说家需要极力书写的一个重要场域,他宣称:“我们在何种程度上向机器下放道德决定是一个非常尴尬而有趣的旅程。遗憾的是,我已经70岁了,不能见证更多的故事内容。小说家可以在我们与机器在道德领域的互动这个领域大显身手”(Aron 2019:43)。

无论是阿西莫夫、图灵还是麦克尤恩,他们关于人类与机器人之间关系的观点至多只代表了部分人类的立场和态度,我们不禁要反过来假想这样的问题:机器人对其与人类之间的关系又持有怎样的态度与立场?通过改写图灵测试和叙述亚当之死,麦克尤恩旨在强调:无论未来的人工智能如何发展,机器人如何得到改进,其在道德领域难以与人类相媲美,毋宁说取代人类。

5. 结语

论及人工智能,尤其是人类与机器人共处的话

题,麦克尤恩认为,“我们将不可避免地面对并阐明三个问题:我们是谁?我们是什么?我们想要什么?”(麦克尤恩2019:67)尽管麦克尤恩本人并未直接回答上述问题,但他在小说《像我这样的机器》中通过引入“人造人”这个话题,有意识地改写图灵测试,将“机器能否思考”这个问题延伸为“机器能否说谎”,并围绕机器人亚当的情感变化及其毁灭等事件,与人工智能之父图灵展开跨时空的对话,触及了机器人之于人类道德生活的介入这一严肃命题。正是在与图灵的对话中,麦克尤恩重审了人类与机器之间的伦理关系,进而提出在人工智能时代,当我们面对比人类更像人类的机器人时,如何可以成为更好的人这一值得深思的问题。就此而言,我们或许可以从约翰·弗兰克·韦弗的话中得到些许启发,即“如果我们把机器人当成人类一样看待,它们也会使我们更具人性”(韦弗2018:233)。

参考文献:

- [1]Aron, J. & I. McEwan. It the 1980s, AI rising and Alan Turing is alive[J]. *New Scientist*, 2019(3226): 42-43.
- [2]McEwan, Ian. *Machines Like Me*[M]. London: Jonathan Cape, 2019.
- [3]Miller, Stuart. Q & A: Ian McEwan on how ‘*Machines Like Me*’ reveals the dark side of artificial intelligence[N]. *Los Angeles Times*, 2019-04-25.
- [4]Theroux, M. *Machines Like Me* by Ian McEwan

Review——Intelligent mischief[N]. *The Guardian*, 2019-04-11.

[5]Turing, A. M. Computing machinery and intelligence[J]. *Mind*, 1950(236): 433-460.

[6]Russel, Stuart J. & P. Norvig. 人工智能:一种现代的方法[M]. 殷建平,等,译.北京:清华大学出版社,2013.

[7]艾萨克·阿西莫夫. 机器人短篇全集[M]. 叶李华,译.南京:江苏凤凰出版社,2014.

[8]巴里·库珀,安德鲁·霍奇斯. 永恒的图灵:20位科学家对图灵思想的解构与超越[C]. 堵丁柱,高晓汾,等,译.北京:机械工业出版社,2018.

[9]J.M.E. 海兰. 被遗忘的图灵[A]. 巴里·库珀,安德鲁·霍奇斯,编. 永恒的图灵:20位科学家对图灵思想的解构与超越[C]. 北京:机械工业出版社,2018:25-38.

[10]安德鲁·霍奇斯. 艾伦·图灵传:如谜的解谜者[M]. 孙天齐,译.长沙:湖南科学技术出版社,2017.

[11]伊恩·麦克尤恩. 如果有一天“人造人”写出了小说——麦克尤恩人民大学演讲[J]. 读写月报,2019(Z2):64-67.

[12]尚必武. 科学选择与伦理选择的冲突:麦克尤恩《像我这样的机器》中的人工智能与脑文本[J]. 外国文学研究,2019(5):61-74.

[13]史蒂芬·卢奇,丹尼·科佩克. 人工智能(第2版)[M]. 林赐,译.北京:人民邮电出版社,2018.

[14]瑞恩·卡洛,等. 人工智能与法律的对话[M]. 陈吉栋,等,译.上海:上海人民出版社,2018.

[15]约翰·弗兰克·韦弗. 机器人是人吗? [M]. 刘海安,等,译.上海:上海人民出版社,2018.

[16]乌戈·帕加罗. 谁为机器人的行为负责? [M]. 张卉林,王黎黎,译.上海:上海人民出版社,2018.

From Alan Turing to Ian McEwan: Artificial Intelligence ,

Lies and Ethics in *Machines Like Me*

Shang Biwu

Abstract: Taking *Machines Like Me* as its point of departure, this paper intends to examine Ian McEwan’s rejoinder to Alan Turing across time. By focusing on the machine Adam’s engagements with human life, it examines McEwan’s re-writing of Turing Test, in which he changes the question "Can machines think?" to "Can machines lie?" Their different responses to Miranda’s lie about her sexual attack reveal Charlie and Adam’s competing ethical positions and attitudes. In contrast to Isaac Asimov’s "Three Laws of Robotics", McEwan proposes the question about human responsibility for machines through Turing’s critique of Charlie’s hammering Adam to death.

Key words: Ian McEwan; Turing Test; artificial intelligence; lies; ethics