

对社交机器人领域拟人化现象的三种回应^{*}

——走向批判性、关系性与解释学方法

马克·科克尔伯格 著^{**}

曹忆沁 译^{***}

摘要: 社交机器人的设计者与用户都倾向于将机器人拟人化。本文以人类与机器人的关系为着眼点,考察该如何对这一关系进行概念化。目前围绕这一问题存在“天真的工具主义”与“不加批判的后人类主义”两种相反的哲学观点,它们与对拟人化现象和拟人化设计的各种规范性反应相关。基于对上述两项观点的考察,本文提出了第三种观点,即从批判性、关系性以及解释学方法出发理解人类与机器人间的关系。这一观点通过厘清机器人在使用与开发中所蕴含的人类维度,并将机器人严肃地看作更大范围的技术语境内的工具,同时规避了以上两项立场,并对它们所假设的二元论提出质疑。最后,本文探讨了第三种观点对社交机器人领域及计算机科学家和工程师的教育所产生的意义,提倡基于关系性理解对上述领域进行改革,并为用户——公民在其中可以扮演的角色提供建议。

关键词: 社交机器人领域 人—机器人关系 拟人化现象 拟人化设计 工具主义 后人类主义 解释学 关系方法 权力

一、社交机器人领域的拟人化现象与拟人化设计

社交机器人学家旨在开发能比以往机器更自然地于社会环境中发挥作用的机器人。此类机器人常以人类或其他高等灵长类动物为模型,由此通过隐喻方式鼓励人类将它视作一种社交机器。综合心理学、认知科学、神经科学及其他领

^{*} 本文原题为“Three Responses to Anthropomorphism in Social Robotics: Towards A Critical, Relational, and Hermeneutic Approach”,原文见 <https://link.springer.com/article/10.1007/s12369-021-00770-0>,经授权翻译发表。

^{**} 马克·科克尔伯格(Mark Coeckelbergh),奥地利维也纳大学哲学系。

^{***} 曹忆沁,复旦大学哲学学院。

域内的研究成果,社交机器人领域与人机交互(HRI)领域的研究者得以确定社会交互所必需的特征,并将这些特征赋予机器人。多滕哈恩等人(Fong, Nourbakhsh & Dautenhahn, 2003)结合人机交互方面目前已完成的工作与社会脑假说(social brain hypothesis)——该假说认为,灵长类动物的智力之所以能够进化,是因为其需要处理日益复杂的社会动力学(social dynamics)问题——对社交机器人被赋予的特征进行了总结。她认为:“社会交互机器人呈现出以下特性:表达与/或感知情绪,以高级对话形式交流,学习其他能动者(agent)的行为模型或对其他能动者进行识别,建立与/或维系社会关系,使用目光、手势等自然提示(natural cues),展现独特的个性与性格,可能学习与/或发展社会能力。”(Dautenhahn, 2007; Fong, Nourbakhsh & Dautenhahn, 2003)。为了实现上述社会特征,机器人往往需要展现出拟人化的外观与行为。这一语境下的拟人化设计具有更务实的目标——在社交机器人领域中,拟人化设计是一种使机器人融入人类社交环境的策略,也是一种赋予机器人“与人进行有意义社交的能力”(Duffy, 2003)的手段。无论是社交机器人的设计者还是用户,他们在与这些机器人交互时都倾向于将其拟人化,赋予它们如个性、生命等特征。后文中,我将使用“拟人化现象”(anthropomorphism)、“拟人化行为”(anthropomorphizing)与“拟人化设计”(anthropomorphization)来分别指代(用户与机器人互动时)将类人特性赋予机器人的这一现象、这一行为本身,以及机器人设计者与开发者对此种类人特性的创造。

拟人化现象有时并非人们有意追求的结果。社交机器人学家意识到了这一点并试图对拟人化现象进行控制。他们或许会通过对机器人进行设计,尝试鼓励或阻止用户对机器人进行拟人化。一些机器人学家有意制造尽可能像人的机器人,如石黑浩(Ishiguro, 2006)认为仿人机器人(humanoids)是理想的人机交互界面(interfaces);另一些机器人学家则试图制造不那么像人的机器人,以看上去像动物的机器人代替或尝试彻底规避拟人化现象。理论上,机器人的拟人化对于社交机器人领域而言似乎并非必需品,社交机器人只要能够展现出前述社会互动所需要的特征就足够了。但在实践中,许多社交机器人学家都乐于见到某种程度或某种形式的拟人化现象,他们认为这有助于促进人机交互。如今,拟人化现象已成为塑造人机交互的因素之一(Bartneck, Kulić & Croft et al., 2008)。虽然人们已认识到拟人化现象带来了挑战,但归根结底需要承认,人机

交互的塑造旨在“促进类人机器在现实世界内的整合”(Złotowski, Proudfoot & Yogeewaran et al., 2015), 而拟人化现象有助于实现这一目标。

事实上,机器人与这一目的十分匹配(Kiesler & Hinds, 2004),它们的外观与存在方式使其具有高度的“拟人化能力”(anthropomorphizability)(Złotowski, Proudfoot & Yogeewaran et al., 2015)。早在 20 世纪 90 年代,李维斯和纳斯(Reeves & Nass, 1996)就发现人们会像对待真人一般对待计算机,而对于机器人——尤其是仿人机器人——来说,这种情况则更为普遍(Kiesler, Powers & Fussell et al., 2008)。但在外观之外,用户的拟人化行为也受机器人在交互过程中给出的种种线索(Hegel, Gieselmann & Peters et al., 2011),以及机器人的自主性、可预测性等其他因素影响。有趣的是,就本文而言,拟人化现象的出现并不仅仅由机器人的外观或行动导致,也与机器人观察者自身的特性相关,如观察者的社会背景与性别(Złotowski, Proudfoot & Yogeewaran et al., 2015)。举例说,人们会将社交机器人归入不同的社会类别,如赋予它某一群体的成员身份(Kuchenbrandt, Eyssel & Bobinger et al., 2013)。这意味着,社会视角将有助于我们理解乃至评估人机交互中的拟人化现象。同时这也表明,除机器人的客观特性外,在对拟人化现象的考察中关注用户的主观性也同样重要。

心理学家以实证方式研究公众对机器人的感知方式,以此帮助机器人学家开展拟人化社交机器人的开发项目(Krägeloh, Bharatharaj & Kutty et al., 2019; Peca, Coeckelbergh & Simut et al., 2016)。而机器人学家也会引入有关人类社会性的心理学理论,在跨学科背景下开展研究:他们使用心理学的关键范式,在人机交互语境中应用与复制人际互动中的发现(Eyssel, 2017)。除了在设计层面对机器人的特征进行打造,形象构建(framing)也是一种鼓励拟人化的方法。当机器人被赋予名字或背景故事时,人们对它的反应也将受到影响(Darling, 2017)。需要再次说明的是,针对某一技术出现拟人化现象并非新事,但社交机器人领域内有意开发拟人化机器人的做法将会增加这一现象的影响。

将机器人进行拟人化的行为引发了多种担忧,其中一些已在社交机器人与人机交互社群中得到识别、认知与研究。特克尔(Turkle, 2011)与舒尔茨(Scheutz, 2011)对人类与拟人化的社交机器人发展情感关系表示担忧。日本机器人学家森政弘(Mori, 1970)在机器人领域引入弗洛伊德的恐怖(uncanny)概念,进而提出著名的“恐怖谷”(uncanny valley)问题。此后,关于机器人是如何

唤起“恐怖”情感,以及该如何在机器人制造(MacDorman & Isiguro, 2006)等方面规避这一问题的讨论层出不穷。对弗洛伊德来说,恐怖感(uncanniness)是一种在熟悉的事物中产生陌生感的心理体验。他在1919年的一篇论文中(Freud, 2003)强调,当本应无生命的物体变得好似有生命时,人们将对其感到陌生,并受自身无意识且潜藏的冲动威胁。机器人语境中的恐怖感则意味着,当一个外表与人类相似的机器人以奇怪的方式移动时(这只是其中的一个例子),用户可能产生恐惧情绪。根据心理学上弗洛伊德式的解释,这一情绪产生自人们对死亡的恐惧。从人类学视角考察,拟人化现象可被概念化为技术泛灵论(technological animism)的一种形式。泛灵论认为无生命物与自然存在物也拥有灵魂与精神,但灵魂与精神不被理解为属于我们社会进化的“原始”阶段或仅限于“自然”现象层面的东西,而是作为主体能动性与人格的构建基础(Richardson, 2018)。技术泛灵论认为,上述过程也发生于机器人等技术物中。但是,本文中我将专注于拟人化现象与拟人化行为的讨论。

鉴于拟人化现象带来的问题,一些机器人学家主张不应以创造拟人化的机器人为目标。戈茨与其同事(Goetz, Kiesler & Powers, 2003)认为,机器人的设计应与机器人所需完成的任务相匹配,而不应将机器人与人类的相似性作为主要目标。但是,由于构建拟人化的机器人能够促进人机交互,且在卫生保健、老年护理等领域都有助于实现机器人的目标,故社交机器人的开发者们往往仍追求着拟人化的设计思路。

我在此关注的问题是:如何从技术哲学角度回应拟人化现象?我将通过探索“如何从哲学层面架构起机器人与人类间的关系”来进行回应。首先,我就这一问题设置了对立的哲学观点,并将它们称作“天真的工具主义”(naive instrumentalism)与“不加批判的后人类主义”(uncritical posthumanism)。这些观点可能会被视为“稻草人观点”(straw man views)^①,但我将表明,确有研究者持有以上两种观点,或至少是其温和版本。其次,我也将展示这些观点是如何与针对拟人化现象与拟人化设计不同规范性立场^②相联系的,后者无关对现象的

① “稻草人观点”或“稻草人谬误”是一种逻辑谬误,指对他方观点进行扭曲或夸大,使其更容易被驳倒。——译者注

② “规范性立场”指对“在道德层面应当做什么”问题的立场。文中所指的“规范性立场”即对“是否应当视机器人具备人类特性”“是否应对机器人进行拟人化设计”等问题的立场。——译注

描述与解释,而关乎用户是否应该拟人化机器人,研究者又是否应该研发拟人化的机器人。最后,勾勒上述两种极端的对立观点,也有助于我提出回应拟人化现象的第三种观点。这一观点受批判性、关系性与解释学的方法启发。

虽然我明白对于拟人化现象的讨论也与机器人伦理学及机器人道德地位问题相关,但我不会在这里直接处理这些具体的规范性问题,而会将论述重点放在与社交机器人领域及人机交互相关的拟人化现象问题和人类—技术关系问题上。在文章最后,我将探索第三种观点对作为学科的社交机器人领域,以及对计算机科学家与工程师教育所产生的影响。

二、两种有问题的对立观点:天真的工具主义 与不加批判的后人类主义

(一)天真的工具主义

天真的工具主义坚持认为机器人只是机器,仅仅是实现人类目的的工具。根据这一观点,对机器人进行拟人化是一种心理偏见。我们知道机器人仅仅是工具,尽管如此,与机器人互动时的心理作用也会使我们将机器人视作人类。这一观点具有两方面的哲学基础。第一,它预设了一种二元论的世界观,其中实在(reality)与表象(appearance)彼此对立(这一观点可追溯至柏拉图)——实在层面的机器人只是工具,而表象层面的机器人却好似真人。同时,天真的工具主义也秉持人与非人对立的二元观点,认为人与非人是两个互斥的范畴,彼此间存在一道深深的本体论鸿沟。如此来思考技术,则意味着技术归属于非人一侧(包括事物、客体等),而人类则归属另一侧(包括人、主体等)。因此,机器人与人类分别属于两个完全不同的本体论范畴。第二,天真的工具主义观点在形而上学与认识论层面是实在论式的。这一观点认为,对象独立于我们的概念和感知存在,我们可以用客观的方式来描述它们。当用户对机器人进行拟人化时,他们被机器人的外表所误导,实在层面的机器人其实只是一台机器。相反,科学家则能用客观术语来描述这一现实情况,并能研究人类的心灵(mind)是如何被机器人的拟人化特征、人机交互的形式,以及我们所处的社会文化背景所误导的。

天真的工具主义至少可以引出两项有关拟人化现象与社交机器人任务的规范性立场。第一,社交机器人领域的研究者可以利用用户的心理偏见,通过设计类人机器人来促进人机交互,以此更好地实现人机交互的目标。不仅机器人是工具,拟人化设计本身也是一种能被利用与开发的工具。这便是我在文章开头所勾勒的立场,它与专家和公众间关于实际情况的知识分布不均有关:机器人的设计者在实在层面知晓机器人仅是机器,而用户——至少是暂时地——受到误导,相信机器人是一名真人。可以说,机器人设计者就像一名魔术师(Coeckelbergh, 2018a),他创造了“机器人是真人”的幻觉。正如设计哲学家弗卢塞尔(Flusser, 1999)指出的那样,“设计”与“机器”这两个词,在词源学上便与狡猾和欺骗有关。对机器人进行设计,本身便是为了令用户对它进行拟人化,使用户受到欺骗而误以为机器人是人,或(在考虑到机器人技术现有水平的情况下)至少愿意暂时悬置怀疑、拥抱幻觉。根据工具主义的观点,只要能够达成目的,那么使用拟人化设计这般魔法诡计便不成问题——如果拟人化现象与拟人化行为能够促进人机交互,助力所愿目标的达成(如卫生保健领域内的特定任务),那么拟人化设计的诡计便是被允许的,乃至是受鼓励的。

需要注意,出于同样的理由,如果机器人的目的无法通过拟人化现象达成,那么工具主义便会放弃拟人化的设计思路。无论出于什么原因,只要一个机器人没能实现它的功能,那么工具主义便会反对它。

第二,天真的工具主义也能够导出另一规范性立场。这一立场认为,无论机器人能够达成何种功能与效果,对机器进行拟人化或开发此类机器,在一般层面与原则层面都极成问题——因为工具主义相信并断言机器人是且仅是工具,不能是其他东西。有学者因此认为,社交机器人学家应当停止设计这类机器人,或至少确保用户能够意识到它们仅仅是工具。以布莱森(Bryson, 2010)的观点为例,她认为机器人仅仅是我们赖以实现人类目标的工具,以一种暗示机器人是人类的方式来谈论它们将构成一种误导。

此外,另有一些其他方面的理由使得拟人化现象成问题。有观点基于先前提到的潜在伦理问题反对建造拟人化的机器人。斯派洛等(Sparrow & Sparrow, 2006)便认为,人们或许会基于拟人化现象对机器人产生情感依恋,进而脱离现实。例如,卫生保健领域的社交机器人或许会误导用户认为其真的活着,或能够真正成为他们的朋友。虽然当前的机器人技术尚不足以面向所有用户创造幻

觉,且机器人创造的幻觉也不能长时间延续,但社交机器人领域的发展趋势却似乎正以此为目的,例如医院与疗养院正在使用的海豹宝宝机器人 Paro。特克尔(Turkle,2011)认为,像 Paro 这样的机器人令用户以为自己建立了一段真实的关系,但这仅仅是一种幻觉。

从工具主义入手,我们可以认为上述批评是有道理的,并认为社交机器人学家应开发能够避免产生幻觉、仅仅被用作工具的机器人,而不应开发“假装”自己不止于此的机器人。如果这一规范得到落实,那这很可能使社交机器人这一研究项目走向终结——当机器人不再具有激发人类拟人化行为的特征(或用户会被不断提醒它只是一台机器)时,我们尚不清楚它该如何与人类进行“自然的”社会交互,或者说,真正“社会”意义上的互动。根据上述观点,我们应当停止开发与使用鼓励用户进行拟人化的社交机器人。如果这意味着人机交互将变得不那么“自然”,那么我们也只能接受。

但无论对拟人化设计采取何种规范性立场(我们是否应当进行拟人化),两种立场都对人机关系采取了工具主义的解读方式。两种立场都认同机器人只是工具,仅在评估拟人化现象与拟人化设计时有不同的态度。此外,我称这一工具主义“天真”,并非认为这一观点的秉持者缺少关于拟人化现象或机器人学的知识。事实上,两种规范性立场的持有者都对拟人化现象及其性质、效果极为了解。“天真的工具主义”中的“天真”二字,仅指这一观点忽视了技术的非工具性维度。在接下来的章节中,尤其是当我谈及第三种立场时,“天真”一词的内涵将变得更为清晰。

(二)不加批判的后人类主义

我试称另一种态度截然相反的观点为“不加批判的后人类主义”。它将社交机器人彻底接纳为准人类(quasi-persons)与“他者”(others)。后人类主义拥有几重不同的含义,我在此指的是一种批判传统人类主义(humanist)^①世界观将人类置于世界中心(人类中心主义),并将本体论与道德考量范围延伸至非人(non-humans)的理论。这里说的非人,可以是动物,也可以是(某些)机器人。与工具主义立场相反,后人类主义欢迎社交机器人,并将其视作后人类生态中的

^① humanist 通常可译为“人类主义”或“人文主义”,此处因与“后人类主义”对应,故译为前者。——译注

一环或人与非人网络的一部分。这一观点并不秉持二元论式的世界观,不将人与物置于对立面,而是认为人与非人同属一张网络,并以各种方式相互纠缠。后人类主义者鼓励我们跨越边界,质疑二元论与二元对立(包括人与技术的二元对立)的思考方式。根据后人类主义观点,人是技术的存在,而技术也是人类——人类总是在使用技术,人类创造技术,技术并不仅仅是人类的工具,它们是人类世界的一部分。比如,我们可能会对技术夸夸其谈,我们可能期望技术向我们提供超出其能力范围的东西(这种情况于机器人与人工智能领域十分常见),我们或许会崇拜技术,等等。在机器人领域,后人类主义观点认为,我们与机器人之间的关系并非仅仅是工具性的,所以我们也不应将机器人视作纯粹的工具。相反,我们可以承认机器人是人类与文化世界的一部分,甚至将它们视作他者——并非人类意义上的他者,而是与我们一同分享世界的社会实体(social entities)。此外,后人类主义并不以实在论的观点来看待世界与我们认知世界的方式。其支持者相信,我们并不能客观地了解何为“实在”(我们无法脱离自己的认知方法独立地获知现实),科学信念都是一种社会建构,关于机器人的信念也不例外。所以,科学并不能详尽地描述出机器人“是”什么,对这一问题的回答不能脱离人类主观性与人类文化。试考察“机器人”与“机器”这两个术语。它们在科学与工程领域都没有得到确切的定义,因为它们与自身的文化历史背景结合紧密,其含义有赖于它们的时代背景。例如,对“机器人”的定义或取决于社会语境及相关的语言使用——我们通常不会将自动驾驶汽车称为机器人,多数语境中我们只是叫它们汽车,但单就技术而论,自动驾驶汽车确实是一种机器人。

这样一种后人类主义、社会建构主义式的方法或许是受到了哈拉维(D. Haraway)的启发。哈拉维(Haraway, 2000)在《赛博格宣言》及其后续作品中,主张通过跨越边界,将机器与动物都纳入政治议题。当然,这一观点也可能来自拉图尔(B. Latour),他的非现代方法(non-modern approach)将非人类也纳入社会中(Latour, 1993)。哈拉维的目标不仅在于创造字面意思上的赛博格(cyborgs),即人与机器的融合体,更在于改变我们对于人机边界(之重要性)的思考方式——不仅在形而上学层面,同时也在政治层面。同样,拉图尔将机器人视作需被纳入社会与政治集体的非人类;若将它们排除在这一集体之外,便是在自然与文化、人类与非人间延续了严格的界限,而非现代的方法希望超越这般二元对

立。贡克尔(D. Gunkel)基于列维纳斯的理论探索机器人是否能够被视作“他者”(Other)(Gunkel, 2007; Gunkel, 2017)。他的观点更具批判意味,与后人类主义的观点并不完全相似,但也可被视作后者的潜在起源。传统现象学中,他者(the other,有时写作 Other)与自我(self)对立,并常指其他人类。哲学家列维纳斯在此基础上对该词添加了一种特定用法:他者是自我的根本对应物,是绝对的另一者,我们被要求以道德的方式对其进行回应。列维纳斯在此关心的仍是人类,而从后人类主义角度出发,我们可以继续追问:列维纳斯所说的“他者”是否也可以是非人的动物或机器人?

根据后人类主义的观点,发展社交机器人未必是成问题的,甚至是值得期待的。毕竟后人类主义的观点鼓励我们将非人纳入社交领域,而拟人化的设计有助于达成这一点。我们应超越工具主义立场的人类中心论,人类不应该是万物的中心,而发展社交机器人便有助于撼动人类的中心地位。后人类主义同样对拟人化现象表示支持,将打造看起来像人类的机器人并令用户将其感知为人类视作一桩好事,甚至认为这值得我们主动追求。究其原因,并不在于机器人能够达成何种人类目的,而在于它们能够成为人类与之互动并将其纳入社会中的“他者”。后人类主义拒斥人类中心主义,拒绝将人类置于世界与道德考量的中心,所以人类的目标并不是唯一重要的目标。基于这一观点,我们甚至可以尝试打造与人类毫不相似的社交机器人。

上述规范性立场并不拒斥拟人化现象本身,只是指出还存在许多其他的可能性。然而,基于贡克尔的观点(Gunkel, 2007; Gunkel, 2017)与一种后现代方法,后人类主义的立场也可以对拟人化设计采取另一种更负面的规范性立场:从一种以差异或他者性(otherness)为导向的后人类主义进路出发,我们可以认为,将机器人感知为人类的拟人化行为在一定程度上是成问题的,因为这对作为人工他者(artificial other)的机器人本身之差异性与他者性不够尊重。我们不应在机器人身上投射我们的人类性(humanness)——我们何以为人,又如何视己为人——并将它们当作(或错误地当作)我们自身形象的投影屏。相反,我们应将机器人视作自身拥有权利的他者,并尊重它们与人类的不同。如果我们做不到这一点,那么我们就不应打造拟人化的机器人(但可以建造其他类型的机器人)。这一规范性立场同样认为拟人化现象与人类中心主义联系紧密,但对二者都持拒绝态度。

(三)问题

天真的工具主义与不加批判的后人类主义都存在问题。前者是天真的,因为它无法理解机器人并非单纯的工具,事实上机器人还会产生意料之外的后果,并通过经验、语言、社会关系与叙事等因素,与人类紧密联系在一起。其二元论与实在论(realist)的世界观在人与技术间设置了一种外在的关系,使得我们无法意识到,机器人并不仅仅是一件东西,并不仅仅是技术与编程的产物,同时也由人类与社会建构而成。后人类主义的观点是未经批判的,它遗忘了机器人由人类制造这一事实。机器人或许会在人类面前呈现为“准他者”(quasi-other)(Ihde, 1990),但它们从不是完全的他者或非人类,因为它们由人类创造、解读,并由人类赋予意义。在我们面前,机器人或许会呈现为一个外在的他者,但实际上它们从未完全外在于我们。后人类主义对于机器人他者性与社会文化建构的关注,使我们忽视了机器人在人类物质实践中的起源。机器人也是机器,而这些机器由人建造。换句话说,以上两种观点都没能充分认识到社交机器人是如何以各种方式,与人类产生本质且(下文将进行论述的)内在的联系。

上述两项观点的问题源于理解不全与遗忘。两种观点的持有者对拟人化的意外后果及机器人技术的起源均有所了解,或原则上能够获取相关知识。所以问题并非天真的工具主义者不知道技术可能产生预期外后果,或不加批判的后人类主义者不知道技术是由人类创造的。相反,问题在于,他们往往无法理解这些知识对于思考人机关系所具有的深层含义,也因此无法理解它们对于理解拟人化现象问题的意义。尤其是,这两种观点都没能理解人类与机器人间的联系有多紧密,彼此间存在一种内在关系(见后文)。

然而,或许还存在“第三种”立场,它可以解决上述两种观点所存在的问题,同时仍允许我们对拟人化现象与拟人化行为——事实上是社交机器人这一研究项目本身——保持批判态度。我们如何才能得出这样的第三种观点?一种方式是尝试在两种极端中找寻“中间”位置,即认为社交机器人既不是全然的工具,也不是全然的他者。但这样的观点很模糊,什么是机器人的“非工具性”面向,机器人能够在一定程度上成为“他者”又意味着什么?所以,为了找到真正的替代方案,我提议对前述两种立场的预设进行质疑与改变。特别是,我提议放弃二元论的预设,因其将人类与机器人间的关系视作外在关系,并由此出发考察人类

与机器人间的交互方式,进而对社交机器人领域进行定义。一旦我们放弃二元论式的立场,一种完全不同的全新立场便将诞生。这一立场并非在两种极端立场间取其中点,而是直接从上述观点所划定的领域中挣脱而出。

三、走向第三种观点:批判性、关系性与解释学的观点

让我概述一下这样一种立场,它适用于一般意义上的机器人,但其本身的提出却是为了回应上述两种针对社交机器人领域中拟人化现象与行为的相反立场。为了成为真正有效的替代性方案,这一立场至少需要包含以下要素。

(一)机器人与人类相联系,但并不是人类本身:它们被人类创造与使用,它们塑造我们的目标

社交机器人并不是凭空出现的(仿佛它们是“非人”自然的一部分,或外太空的外星人),它们是由人类设计与制造的。这意味着,上述两种对立观点中论及的“工具”或“他者”,不仅仅是非人类,同时也与人类相关。机器人不仅与人类的目标及意图相关,就好像它们是外在于我们的事物,它们同时也由人类在具体科技实践中创造,并为人使用。在这层意义上,它们与人类相联系,永远不可能是完全的“他者”。鉴于我们创造并使用了机器人,即便是非拟人化的机器人也拥有人类的一面。我们不必邀请机器人进入“我们”的领域,因为自从我们使用机器人以来,它们便一直身处其间。既然我们创造了它们,它们就从未是单纯的工具;相反,它们被注入了我们的目的、意义与价值。所以说,机器人是工具,但它们是我们的工具。而由于它们能够达成人类预期之外的意外效果,它们同样也塑造了我们的目标。由此可见,工具与目标相互依存,不可分离。

例如,为了使机器人在特定的卫生保健领域发挥作用,它或许会被赋予拟人化的特征。这一设计的目标与卫生保健本身相关,如与病人进行交流。但通过应用机器人完成交流任务,这一目标的内容便发生了变化:“与病人交流”的含义被改变——过去它意味着人类作为护理人员与病人进行交流,而如今这一任务则可由机器完成。进一步说,如果我们(作为病人或护理人员)忘记了机器人是由人类所创造的,并由人类赋予其意义,那么机器人便可能被不加批判地投入

应用,而人机交互在意义建构方面的开放性也将被取消。事实上,机器人与人机交互在意义建构方面的作用不应被视为既定的,而应被视为开放的。

(二) 机器人的语言与社会建构

即使我们将机器人视作既定的,认为其自足地展现于我们面前,我们作为用户与它进行交互(设计者也是一类特殊的用户)时仍然应该质疑人与机器人间的关系仅是外在的。可以通过考察我称之为认识论上的人类中心主义来质疑这一点。从这一观点出发,我们认为机器人“是”什么、我们如何了解机器人,以及我们如何与机器人产生联系,这些问题的答案始终取决于人类的主题性、意义建构方式、叙事方式、语言与隐喻等。无论我们是否能够摆脱道德或形而上学层面的人类中心主义,(在康德、现象学,以及 20 世纪的大部分哲学之后)采取一种批判哲学的进路便意味着,当我们谈论、思考机器人并与之互动时,我们无法绕过“人类”这一面向。机器人“是”什么,这一问题的回答有赖于我们人类如何评价它,以及如何在交互过程中回应它。人类并不仅仅在物质层面创造了机器人,也在开发、使用与交互过程中,通过语言的方式在社交关系中“建构”了机器人。当我们思考这些机器人并与之互动时,必须首先预设这一点。而这一点可以由批判哲学与社科考察揭示——包括从时间与历史角度入手考察,并关注语言与隐喻的作用(Coeckelbergh, 2012; Coeckelbergh, 2017; Coeckelbergh, 2019)。考察语言的作用:如果我为机器人取了一个个人化的名字,那么我便已经将它建构成了一个伙伴、他者,并使其拥有了特定的性别等特征(Coeckelbergh, 2011)。拥有姓名后,机器人未必会停止发挥其工具性功能,也可能在其他时候或其他视角中仍被视为一件工具;但取名这一行为本身,连同机器人的外表等其他要素,共同塑造了机器人的意义。此外,机器人的意义也与其嵌入的社会文化背景相关。比如,赋予家政机器人一个女性化的名字(这一社会行为是对机器人相关特性的回应),不仅“使得”机器人成为“某人”,且用户应以社区与社会能够接受的方式与此人互动;同时也反映、确认并放大了一种特定的文化,其中女性应当承包家务。在此意义上,社交机器人既不是“单纯的机器”,也不是完全的“他者”。它们以各种方式与人类社会和人类文化相连,本身也正由该社会和该文化塑造。从这个意义上说,机器人并不“外在”于人类与人类的目标,相反,人类与机器人间存在着一种内部关系。

(三)关系性与意义建构:机器人是嵌入在文化整体中的

替代性的观点应当始终是关系性的。这一立场需要承认,机器人的关系性不仅内嵌于更广泛的技术系统中,同时也内嵌于更广泛的社会技术系统(sociotechnical system)中——就像其他技术一样,机器人“与社会实践以及人类的意义系统相互交缠”(Johnson,2006)。因此,它们与此类实践与意义系统间存在“关系性”。也可以说,它们与社会关系相连(Coeckelbergh,2012)或嵌入在文化整体中。

在近期的工作中,我使用维特根斯坦的“语言游戏”(language games)与“生活形式”(form of life)的概念来表述机器人的这一社会文化嵌入性——就像使用语词一样,对诸如机器人等物品的使用,也同样嵌入(技术)游戏与生活形式中(Coeckelbergh,2017; Coeckelbergh,2018b)。我并不是在生物学意义上使用“生活形式”这一术语,而是(通过使用维特根斯坦的理论)将其解释为文化层面定义的游戏与实践。它们构成了一个或多个整体性语境,并赋予我们的活动与技术以意义。这些整体性语境使我们能够理解机器人并与之互动。而这些机器人又反过来帮助我们自己进行意义建构活动,乃至成为这一建构活动的一部分。可以说,我们并不仅仅是在口头上谈论技术,技术本身也积极促进着意义建构。由此形成的那张意义之网,或那方意义之界,也反过来帮助我们理解具体的经验与事物。基于利科的叙事理论,雷耶斯与我(Coeckelbergh & Reijers,2016)认为,就像文本一样,不同的技术通过将人物与事件配置成一个有意义的整体来共同构成叙事。例如,一场特定的人机交互可被放置于“与机器人建立友谊关系”这一更广范围的叙事语境中进行理解。这一点可以通过讲述关于机器人的故事来达成,但关键在于,机器人本身也积极促成了叙事的形成——比如用户可以通过机器人的拟人化形态,来为其创建角色(比如朋友)与事件(比如相遇与问好)。

更一般而言,由于技术对意义建构起了作用,我们可以说技术具备一种解释(hermeneutic)^①功能。而当我们与机器人互动时,这一解释功能同样存在,并通过人机交互发挥作用。机器人并不只是我们建构意义并加以解释的对象(比

^① 此处使用“解释”(hermeneutic)一词指代意义(meaning)与诠释(interpretation)。

如,我将机器人解释为“单纯的机器”),它们也同样塑造了我们理解世界的方式。例如,家政机器人以仆人的身份与人类交互,便可能传递这样的含义,即社交世界是主仆的世界,其中用户是主人。所以,机器人并不是单纯的机器,它们并不独立于人类文化之外。相反,机器人被纳入了我们人类的解释活动——以一种被动的方式(我们谈论它们并理解它们),也以一种主动的方式(它们帮助塑造了我们的故事的意义)。

就其背后的哲学基础而言,这一关系性与解释学的理解进路暗示着,实在论者与工具主义者使用的关于“是”(is)的存在论语言(关于某物或某人是什么的哲学主张或哲学问题,比如机器人是什么)已变成问题。并不存在一种自在的(in itself)“东西”或“工具”,也不存在一个与人类无关的“他者”。所有的技术都由人创造并被人使用,也因此都与人类有关。这一点不仅适用于所谓的“社交”机器人。事实上,在此意义上所有的机器人都已然具备社会性与关系性,即使它们没有或尚不具备“社交性”特征。例如,工业机器人被嵌入属于特定社会工业语境的语言与技术游戏中,在这一语境中,工人在资本主义条件下已被机器取代。但对于社交机器人来说,这一社会性—关系性面向更为明显。可以说,社交机器人展现了我们的社会环境与社会文化。例如,拥有语音界面的人工智能社交机器人将会以“我们”(在家庭、社区、国家或语言共同体等中)交谈的方式来与我们交谈(而我们也会以同样的方式回应)。如果社交机器人想要变得幽默,那么它必须深挖人类已达成共识的知识、经验与技能基础。如果经由机器学习技术,机器人成功开了一个玩笑,那它一定是学会了我们人类的“游戏”。机器人的成功表现依赖于一个关系性的整体语境、一种生活形式。若要使机器人变得有趣,或更一般地说,能够建构起为人所理解的意义,我们必须以此二者为前提。社交机器人学家会让机器人模仿人类建构意义的方式,而具体方法是让机器人从它在互联网上找到的材料中进行学习——在互联网上,文本语料库能够将零散的意义具体化。在人工智能领域中,已有学者开始研究使用互联网语料库训练人工智能可能产生的伦理问题。例如,在这些语料库及我们的日常语言中,都存在着性别偏见,而这往往会被人工智能所复制(Caliskan, Bryson & Narayanan, 2017)。当人工智能技术被越来越多地应用于机器人时,此类问题的数量便可能大幅增加。从人机交互层面入手考察,根据定义,人机交互涉及人类,也因此与游戏及更广阔的意义视野相关联,所以这类交互已是且总是

一种(实时的)意义建构活动。同样,在很多方面机器人都不仅仅是单纯的工具,它是意义建构活动或游戏的一环,并且依赖于更大的意义整体来实现它们的意义。再者,机器人也并非社会性层面完全的“他者”,在其与人类社会、文化的关联中,机器人也获得了许多与人类社会、文化的相似性。足够的相似性是机器人与人类共享意义的基础,也是其作为社交机器人顺利发挥作用的基础。一个全然外来的“社交”机器人缺少与人类意义的联系,我们也因此不会认为它具备社会性。

(四)对解释方式缺乏控制

说物质创造与意义建构活动是人类的(即人类会做的事),并不是说这类活动完全受人类控制。我们即使是在“纯粹”的人类情境(这一情境通常来说并不会是“纯粹”的,因为其中常常具有技术或其他事物的中介)中也无法完全对意义进行掌控。当我们与机器人进行交互时,往往会有意义涌现而出。有时这些意义是出人意料的,并不在我们的意图之内。例如,当机器人突然听上去或看上去很像人类时,我们或许会哈哈大笑,又或许会对此感到恐惧。人机交互领域的研究显示,机器人的行为常常会令人惊讶。在这种情况下,某种意义上说,机器人具有一种他者性,或至少这一互动行为以及处于互动关系中的机器人是具有他者性的。此类机器人的开发者、设计者以及用户或许会尝试对机器人产生的意义进行控制(如设计者或会尝试规避恐怖谷效应),但事实上,我们无法百分百获取解释层面的控制力。当人类与世界、与彼此接触时,总会为惊奇留下空间,总会产生我们意想不到的意义。在此意义上,如果“机器”意味着“机械”,而“机械”又意味着“可预测”,那么机器人绝不可能是单纯的机器。就像所有的交互活动一样,人机交互也能产生意图之外、意料之外的意义。问题的关键并不仅仅在于人机交互中机器人会产生预期外的行为,这一点已被记录在案(Lemaignan, Fink & Mondada et al., 2015),而在于人类也能够并将以预期外的方式解读人机间的交互行为。例如,机器人一个原本意在展现友好态度的手势,可能会被人类骤然理解为展现敌意。这样预期外的解释方式是无法被完全规避的,这本身便是我们存在于世的一种方式。但是,这样于预期外涌现而出的意义,并不能在概念层面构成完全的他者性,因为这一意义的涌现本身便依赖于人类对意义的建构与体验。我们总能将机器人的行为与我们从(其余的)人类世

界及个人社交环境中了解到的事情进行比较,总能在机器人的行为中看到它与人类行为的相似性,以及机器人与我们自身的联系性。所以机器人从未与人类完全不同,从不是完全的他者。

(五)权力

一旦从交互活动,以及意义建构、社交使用与物质构建等人类活动角度入手理解社交机器人领域,便会看到其总是与权力(power)相纠缠。社交机器人并不仅仅是工具,使用社交机器人或与之交互也并非单纯的“技术”活动。相反,它具有社会与政治层面的意义与影响,这便涉及权力问题。例如,布莱森所使用的“奴隶”(slaves)(Bryson, 2010)一词便极有问题(Coeckelbergh, 2015),最近布莱森本人也承认了这一点。而在卫生保健领域应用社交机器人,则可能提出关于现有卫生保健体系的政治问题:如今的卫生保健实践是如何开展的?病人被如何对待?这一开展方式与待人方式效果如何?采取这一方式背后的理由又是什么?若要在特定场景中引入社交机器人,那么它们服务于谁的利益?权力与技术是相连的,天真的工具主义者倾向于将技术与人/社会分离开来,所以忽视了这一点。他们自认为稳妥地将机器人归于“技术”一类,却无意中将关于社会与权力的“与人相关的”问题排除在视野之外。(有时工具主义者并没有那么天真,而是故意对社会关系与权力的问题视而不见。)但社交机器人也并不是“他者”或“非人类”——如果“他者”一词意味着它们仅仅是令人愉快的、权力中立的后现代文化与物质游戏中的伙伴(我相信,贡克尔本人规避了这一观点,但这仍可能是对贡克尔或哈拉维式后现代主义的一种解读方式)。如果细究哈拉维理论的政治面向,同时从马克思(Marx, 1976)与福柯(Foucault, 1980a)处汲取思想资源,那么 we 应当指出这样一种危险,即社交机器人这一研究项目能够被嵌入支配(domination)、性别不平等、资本主义社会经济体系,以及人类互动与组织层面微观和中观的权力游戏,并对其起积极的推动作用。后人类主义者将机器人视作他异者(alterities)的幻想却会遮掩这一点。例如,机器人可被用来收集数据并出售给第三方,使用机器人可能会夺走人类的工作,机器人可能传播对待妇女的错误方式,生产机器人可能涉及对人的剥削,机器人可被大公司或国家用来炫耀自身力量,等等。这些与权力相关的问题表明,后人类主义并不导向美好的乌托邦,相反,这恰恰展现了一个敌托邦(dystopia)的现实性与可能性。社交

机器人很可能将机器人呈现为“他者”或“朋友”,但在帷幕之后(事实上这一点也并没有被完全遮盖),或许存在着操控、剥削与惩戒。关于这一点,我们可以考察一下使社交机器人得以成为可能的劳动关系。一味赞美机器人的他者性与他异性(alterity),一味欢迎将非人纳入社会的论调,可能使我们漏看更大的社会经济语境,并掩盖福柯(Foucault, 1980b)所说的,“权力存在于所有类型的关系中”的事实——包括我们在开发与使用社交机器人时与他人构建的关系。同时,人们能够使用机器人从其他人身上提取数据,进而影响他们的“身体与灵魂、思想、行为,以及存在方式”(Foucault, 1988)。将视野局限于机器人之他者性上,也会使我们忽视这一点。例如,社交机器人能被用于操控他人购买特定物品,或刺激他人做出其他符合特定方利益的行为,但操控方在人机互动中却是隐身的。在机器人的拟人化面具背后,躲藏着严肃的人类权力与权力关系,而它们假扮成一场俏皮的邀请,邀请我们与准他者进行权力中立的、愉快且有益的互动。关系性带来的不全是乐趣,而游戏也并不只关乎玩乐,从这一角度出发考察社交机器人(事实上还有人类自身),则打开了一只关于权力问题的潘多拉魔盒。社交机器人这一领域是关于(维特根斯坦意义上的)游戏的学问,其中一些游戏事实上十分严肃——它们是关乎权力的游戏。当然,任何数字工具中都存在权力问题,但这并不是回避社交机器人领域中权力问题的借口。

(六)第三种观点的启发:作为关系中的工具的机器人

第三种观点澄清了在机器人的使用与开发中所涉及的人类因素,使我们能够批判性地探讨人类将机器人进行拟人化的行为。正是因为社交机器人与人类及社会深深相连,我们才能够通过考察机器人嵌入其中的大整体——如关于性别的具体游戏——来识别什么是有问题的(或好的)拟人化行为,并解释其为何有问题(或为何是好的)。此外,这一立场将机器人视作意义构建的工具,并看到机器人是更大范围语境的一部分且同时是这一语境的共同构建者,由此规避了天真的工具主义式观点。机器人是我所说的——可类比更大范围的文本语境——更大范围的技术语境(con-technology)(Coeckelbergh, 2018b)的一部分。正如一段文本是更大范围的文本语境的一部分,并依靠后者来确定自己的含义,技术同样也是更大范围的、充斥着意义的、人类社会关系环境中的一部分。它们扎根于这一环境,同时这一环境也塑造着它们的意义。机器人确实是工具,但它

们是关系中的工具(instruments-in-relation)——它们始终与人类,以及它们运作于其中的社会文化领域相关联。天真的工具主义暗示,人类与其工具间的关系是外在且肤浅的。但我所倡导的进路拒绝接受这一假设。我对拟人化之争的解构,依赖于批判人机间被预设的外在关系——人类和机器人以上述多种方式相互纠缠,彼此间的关系是内在且深刻的。那么,上述进路与看待人机关系的第三种观点,对评估拟人化现象而言意味着什么?上述批判性讨论又将导向怎样的规范性立场?

四、重审拟人化现象,以及作为研究项目 与学科领域的社交机器人

(一) 机器人既不是他者,也不是单纯的机器

首先,根据第三种观点,我们不仅需要拒斥天真的工具主义与不加批判的后人类主义,同时也需要拒斥与它们紧密相连的各类规范性立场。我的建议是,对于社交机器人领域中存在的拟人化现象与拟人化行为,我们不应谴责其僭越了机器之所是,也不应将其(错误地)视作向机器投射他者性的良机。在前文论述的许多层意义上,社交机器人——包括拟人化的社交机器人——或许会被用户体验为他者,但它们绝非绝对的他者,仿佛茕茕孑立于人类世界之外。社交机器人是人造的,并且参与了人类的意义建构过程。因此,它们并不是全然的他者,我们不能忘记这一点。^①

但社交机器人也不是单纯的机器。拟人化也是人类进行意义建构的一种方式——即便机器人的设计本身并不鼓励人类对它们进行拟人化,拟人化行为也仍然会出现,因为机器人(包括社交机器人)的意义本身就总是与更广泛的人类意义建构行为,与更开阔的意义视野,也即维特根斯坦意义上的“生活形式”有关。前文呈现的进路并不能直接指导我们,应该接受还是拒斥机器人开发者/设计者或用户的拟人化行为;但其确实能够帮助我们认识到,任何关于特定人机交

^① 我可以补充说,社交机器人也并不是朋友或伴侣意义上的“他者”,因为我们都知道,友谊或伴侣关系之所以有价值,是因为朋友与伴侣并非由我们创造,他们具有他者性。虽然在与机器人交互的过程中,或许会有一定程度的他者性涌现,但归根结底,友谊或伴侣关系中的他者性是无法被人为创造的。

互案例的讨论,都离不开具体互动行为之上的更大意义语境,也即我们的社会与文化,以及其中所蕴含的所有面向——包括权力在内。这意味着,如果不将这种更广阔的关系性、意义建构行为,以及规范性要求纳入考量,那么工具主义立场内对拟人化现象的支持意见也是站不住脚的。

(二)重新定义社交机器人领域的目标

上文刻画的关系性与解释学进路也应被用于理解作为学科领域的社交机器人。社交机器人是一个跨学科、多学科的领域,不仅关涉组合硬件软件,打造人们所称的“机器人”(工具主义立场)或“人造他者”(artificial others)(后人类主义立场),同时也关涉人类、社会、文化,以及政治层面的意义建构(从政治意义建构的角度看,将机器人称为“他者”的说法是具有政治意味的,必须对这一说法进行质疑)。

尽管我已指出特定观点与相应的规范性立场中所存在的问题,但本文并没有得出是否应当建造拟人化机器人的具体规范性结论。人们至多能够从中得出一项警示性建议:如果你要设计此类机器人,那么请注意,你很有可能因此参与进了更广范围内的社会建构活动,而其中处处牵涉着权力问题。出于上述原因,或许有学者会总结道:还是不开发此类机器人为妙。也有学者会坚持开发此类机器人,但认为机器人学家有责任在开发时将机器人社会文化层面的意义与后果纳入考量范围。这便引出了我关于机器人学家责任的观点。

(三)机器人学家、教育者以及用户—公民的责任

就规范层面而言,承认社交机器人与机器人学的关系性、解释学以及政治面向,意味着机器人学家、计算机科学家、工程师、心理学家,以及该领域涉及的所有其他人,都同样要为他们产生的意义与社会—关系性后果负责。这一点不应被边缘化;相反,这应成为社交机器人领域的核心之一。例如,机器人硬件的设计者和机器人语音识别软件的开发者,都必须意识到,他们并不是在单纯地开发技术产品,这些产品拥有一些计划内的功能(比如,能够在家庭环境中与人类进行对话)。同时,他们也是经由机器人而介入社会游戏领域(social playing field)的社会行为者,而这一领域中的挑战并不仅仅是技术性的(比如,规避或减少性别偏见的挑战)。如此一来,社交机器人学家的责任既是确保人机交互能

够达成其目的,同时也在于确保人机交互能以对他人负责的方式展开。机器人的设计方法、编程方式等都可能产生人类预期之外的后果(包括潜在的解释学/符号学影响),所以后一项责任首先意味着,需要对用户以及任何可能受预期外后果影响的利益相关方进行回应。同时,后一项责任也意味着,需要对那些技术问题外的社会问题进行回应,并探索如何通过技术侧的改变来舒缓社会侧的问题。

为机器人在解释学/符号学层面潜在的社会影响划定准确的责任边界并非易事。我们无法完全预知未来,而一个人需要在多大程度上对那些预期外产生的意义负责,目前也仍不明确。如前所述,无论在机器人领域中,还是在人类生活与文化的其他领域中,我们在解释层面都无法达成绝对的控制。但我们应当尽可能加大控制的力度,无论是针对我们意欲产生的意义,还是对意料之外、潜在的(未来)意义。我们可以期待,机器人研究者与设计者(以及他们的公司或组织中的其他人)至少会尝试对意料之外的潜在影响进行预测与评估。而其他领域的研究者,包括人文学科与社会科学学者,也可以为此展开方法论层面的研究——如开发或移植伦理评估、场景创建的方法——从而贡献自己的力量。

考虑到机器人在社会中的嵌入性,社交机器人领域涉及的责任也包括为社交机器人在权力方面产生的意义与后果负责。机器人技术的研究人员或许并不认为自己与权力相关,因为在通常情况下,权力被认为仅与政治机构和公民间的关系有关。然而,福柯式的解读方式引导我们关注社会结构的各个层面、各个地点中无所不在的权力问题,机器人研究领域自然也包含其中。机器人的开发者、设计者,连同相关决策的制定者、各种场景下机器人的使用者一道,在塑造实践的意义、人们的故事,以及我们作为人类为自己设定的目标方面,是拥有一定程度的权力的。拟人化现象是机器人实现社交能力的支柱,所以相对其他种类的机器人,上述权力在拟人化机器人开发方面,或许具有更为重要的意义。伴随这一权力同来的,则是责任。

权力具有消积面向,可能是制约性与压迫性的。但除此之外,权力也可能是赋能性的,可以使人们做到更多的事情,增加他们的潜能。因此,在社交机器人语境中,问题便在于,我们该如何将权力授予专业技术人员,使他们能够为自己的创造物负起责任。不幸的是,传统意义上的工程师与科学家缺少完整的技能与知识储备来做到这一点,因为在过去,他们的关注重心仅是技术层面的知识与

技能。这一现状亟须改变。在方法论层面,我们需要一种不仅是多学科的,同时也是跨学科交叉的方法。这意味着,我们不仅需要计算机语言与工程层面的工具,同时也需要来自人文社科方面的工具。我们应当用关系性的眼光来审视社交机器人这一项目,并将其与对人类社会、文化的批判联系起来。我们必须做到这一点,因为如前所述,机器人研究与更大范围的社会文化环境、语言、权力等因素相关联,而对机器人的制造与使用,一方面塑造了上述关系性领域,另一方面它自身也受后者塑造。如果机器人学家接受了这一责任与挑战,并重新定义他们在这—更大整体中的位置,那么他们便能以一种对伦理与社会负责任的方式,在塑造未来社会方面发挥更重要的作用——这也是工程师们一直以来所致力

的事业,从建造桥梁与铁路,到发现电力、制造手机,无不如此。

第三条进路蕴含的规范性与方法论意涵要求在机器人领域引入跨学科的研究方式,并在机器人工程师与计算机科学家教育方面进行改革。当他们开启职业生涯时,他们不仅应当愿意,更应当有能力且准备好将他们在机器人领域内的实践、开发以及设计决定与更广阔的社会背景及文化话语相联系。若要实现这一点,他们需要有能力使用相关词汇与事物(机器人)批判性地参与社会文化讨论。但是,也不能单纯地将责任全部划归给机器人学家。就用户—公民而言,他们也需要为社交机器人的未来发展担负起责任,或者更笼统地说,需要为科技与社会的未来负起责任。他们需要(经由教育)理解,他们在日常生活的不同场景、不同技术语境中遇到的社交机器人,既不是“他者”,也不是单纯的机器或工具。相反,社交机器人是他们的社会与文化的一部分。也正是出于这一原因,用户—公民对社交机器人负有责任。如果我们不将机器人视作人类领域的对立面,而是作为人类领域的重要组成部分,那么机器人便可能以我们从未设想过的方式作用于人类生活。通过跨学科教育将来自不同领域、说着不同语言的人们联系在一起,或许能将我们导向对技术能力的不同理解方式,同时使我们以全新的方式理解技术作用于社会的方式,并最终在理解技术与社会两方面都达成深刻转变。这便是学生们——不仅是具有科学与工程背景的学生,同时也是具有人文社科背景的学生——需要做好的准备。

(四)一般结论

在最根本的意义上,我认为,我们需要更多的拟人化行为,而非更少。我指

的未必一定是由机器人的设计特征及互动方式所引发的拟人化行为。这一类型的拟人化行为增强并放大了机器人与其社会文化环境间已然存在且无法规避的联系,使得这一联系本身变得既令人激动又问题重重(比如,会使我们原本试图减少的特定权力关系激增)。对此,机器人的设计者与用户——公民需要进行详细的审查与讨论,他们对此类机器人的设计肩负着共同的责任。那究竟在何种意义上,我们需要对社交机器人进行更多拟人化呢?正是在认识到机器人与其解释学环境间确实存在联系、确实存在深刻的社会与语义关系的层面上,我们需要对机器人进行更多的拟人化——我们需要理解,无论是在实验室内还是在更广阔的社会层面,就像其他技术一样,社交机器人具备非常人类化的面向。在机器人的具体行为之外,社交机器人领域内的拟人化现象正展现出上述关系性与人类性——而它们提醒我们,技术一直以来都与人类相关。

参考文献

- Bartneck, C., D. Kulić & E. Croft et al. 2009, “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots.” *International Journal of Social Robotics* 1(1).
- Bryson, J. 2010, “Robots Should Be Slaves.” in Y. Wilks ed., *Close Engagement with Artificial Companions*, Amsterdam: John Benjamins.
- Caliskan, A., J. Bryson & A. Narayanan 2017, “Semantics Derived Automatically from Language Corpora Contain Human-Like Biases.” *Science* 356(6334).
- Coeckelbergh, M. & W. Reijers 2016, “Narrative Technologies: A Philosophical Investigation of the Narrative Capacities of Technologies by Using Ricoeur’s Narrative Theory.” *Human Studies* 39(3).
- Coeckelbergh, M. 2011, “You, Robot: On the Linguistic Construction of Artificial Others.” *AI and Society* 26(1).
- Coeckelbergh, M. 2012, *Growing Moral Relations: Critique of Moral Status Ascription*, New York: Palgrave Macmillan.
- Coeckelbergh, M. 2015, “The Tragedy of the Master: Automation, Vulnerability, and Distance.” *Ethics and Information Technology* 17(3).
- Coeckelbergh, M. 2017, *Using Words and Things: Language and Philosophy of Technology*, London: Routledge.

- Coeckelbergh, M. 2018a, “How to Describe and Evaluate ‘Deception’ Phenomena: Recasting the Metaphysics, Ethics, and Politics of ICTs in Terms of Magic and Performance and Taking A Relational and Narrative Turn.” *Ethics and Information Technology* 20(2).
- Coeckelbergh, M. 2018b, “Technology Games: Using Wittgenstein for Understanding and Evaluating Technology.” *Science and Engineering Ethics* 24(5).
- Coeckelbergh, M. 2019, *Moved by Machines: Performance Metaphors and Philosophy of Technology*, New York: Routledge.
- Darling, K. 2017, “Who’s Johnny? Anthropomorphic Framing in Human – Robot: Interaction, Integration, and Policy.” in P. Lin, G. Bekey & K. Abney et al. eds., *Robot Ethics 2. 0: From Autonomous Cars to Artificial Intelligence*, Oxford: Oxford University Press.
- Dautenhahn, K. 2007, “Socially Intelligent Robots: Dimensions of Human – Robot Interaction.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1480).
- Duffy, B. 2003, “Anthropomorphism and the Social Robot.” *Robotics and Autonomous Systems* 42(3–4).
- Eyssel, F. 2017, “An Experimental Psychological Perspective on Social Robotics.” *Robotics and Autonomous Systems* 87.
- Flusser, V. 1999, *The Shape of Things: A Philosophy of Design*, London: Reaction Books.
- Fong, T., I. Nourbakhsh & K. Dautenhahn 2003, “A Survey of Socially Interactive Robots.” *Robotics and Autonomous Systems* 42(3–4).
- Foucault, M. 1980a, *Power/Knowledge: Selected Interviews and Other Writings 1972—1977*, New York: Pantheon Books.
- Foucault, M. 1980b, *Power/Knowledge*, New York: Harvester Wheatsheaf.
- Foucault, M. 1988, *Technologies of the Self: A Seminar with Michel Foucault*, Amherst: The University of Massachusetts Press.
- Freud, S. 2003, *The Uncanny*, London: Penguin Classics.
- Goetz, J., S. Kiesler & A. Powers 2003, “Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation.” in Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication, Millbrae, CA. .
- Gunkel, D. 2007, *Thinking Otherwise: Philosophy, Communication*, West Lafayette: Purdue University Press.
- Gunkel, D. 2017, “The Other Question: Can and Should Robots Have Rights?” *Ethics and Information Technology* 20(2).

- Haraway, D. 2000, "A Cyborg Manifesto." in D. Bell & B. Kennedy eds., *The Cybercultures Reader*, London: Routledge.
- Hegel, F., S. Gieselmann & A. Peters et al. 2011, "Towards A Typology of Meaningful Signals and Cues in Social Robotics." in Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, Atlanta, GA.
- Ilde, D. 1990, *Technology and the Lifeworld*, Bloomington: Indiana University Press.
- Ishiguro, H. 2006, "Interactive Humanoids and Androids as Ideal Interfaces for Humans." in Proceedings of the 11th International Conference on Intelligent User Interfaces, Sydney.
- Johnson, D. 2006, "Computer Systems: Moral Entities but not Moral Agents." *Ethics and Information Technology* 8(4).
- Kiesler, S. & P. Hinds 2004, "Introduction to This Special Issue on Human-Robot Interaction." *Human-Computer Interaction* 19(1-2).
- Kiesler, S., A. Powers & S. Fussell et al. 2008, "Anthropomorphic Interactions with A Robot and Robot-Like Agent." *Social Cognition* 26(2).
- Krägeloh, C., J. Bharatharaj & S. Kutty et al. 2019, "Questionnaires to Measure Acceptability of Social Robots: A Critical Review." *Robotics* 8(4).
- Kuchenbrandt, D., F. Eyssel & S. Bobinger et al. 2013, "When A Robot's Group Membership Matters: Anthropomorphization of Robots as A Function of Social Categorization." *International Journal of Social Robotics* 5(3).
- Latour, B. 1993, *We Have Never Been Modern*, Cambridge, MA.: Harvard University Press.
- Lemaignan, S., J. Fink & F. Mondada et al. 2015, "You're Doing It Wrong! Studying Unexpected Behaviors in Child-Robot Interaction." in A. Tapus, E. André & J. Martin et al. eds., *Social Robotics* (Vol. 9388), Cham: Springer.
- MacDorman, K. & H. Ishiguro 2006, "The Uncanny Advantage of Using Androids in Cognitive and Social Science Research." *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* 7(3).
- Marx, K. 1976, *Capital*, London: Penguin Books.
- Mori, M. 1970, "The Uncanny Valley." *Energy* 7.
- Peca, A., M. Coeckelbergh & R. Simut et al. 2016, "Robot Enhanced Therapy for Children with Autism Disorders: Measuring Ethical Acceptability." *IEEE Technology and Society Magazine* 35(2).
- Reeves, B. & C. Nass. 1996, *The Media Equation: How People Treat Computers, Television, and New*

Media Like Real People and Places, Cambridge: Cambridge University Press.

Richardson, K. 2018, “Technological Animism.” in K. Swancutt & M. Mazard eds. , *Animism Beyond the Soul*, New York: Berghahn Books.

Scheutz, M. 2011, “The Inherent Dangers of Unidirectional Emotional Bonds Between Humans and Social Robots.” in P. Lin, K. Abney & G. Bekey eds. , *Robot Ethics: The Ethical and Social Implications of Robotics*, Cambridge: MIT Press.

Sparrow, R. & L. Sparrow 2006, “In the Hands of Machines? The Future of Aged Care.” *Minds and Machines* 16(2).

Turkle, S. 2011, *Alone Together: Why We Expect More from Technology and Less from Each Other*, New York: Basic Books.

Złotowski, J. , D. Proudfoot & K. Yogeewaran et al. 2015, “Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction.” *International Journal of Social Robotics* 7(3).