

人工智能与性别伦理：一个新时代

马中红

摘要：传统性别伦理以男性/女性二元对立的系统方式成为建构人类社会的磐石之一，而新科技以及人工智能时代的来临，一方面延续了备受自由主义女性主义和后现代女性主义所批评的男女两性的不平等性别存在及其关系模式，强化了男性主体和女性客体的传统性别伦理，另一方面，随着数字青年文化的兴起及其与新科技的结合，又动摇了传统性别伦理的现实基础，再加之科技女性主义有关赛博格/电子人以及人机共存和人类本体自我消失的论断，进一步消解了传统性别伦理，从而迫使我们去思考建立新型的性别平等伦理、多元和差异互补性别伦理以及公平看待人造物的伦理的可能性。

关键词：人工智能 女性主义 青年数字文化 新性别伦理

作为一种基本的社会共识，性别伦理无疑是维持社会结构稳定和族群传承延续的重要价值方式和动力机制，也是人类思考人性、理解自身和社会交往的主要辨识指标。但是晚近以来，建基于男女二分法、建立在婚姻、家庭、传宗接代等核心概念之上的传统性别伦理已经遭遇到全球范围内女性主义的不断挑战。进入当下，人类开始走进所谓人工智能时代，性别伦理更是面临一系列新技术、新算法及其应用产品的挑战。虚拟人偶、智能语音助手、虚拟伴侣机器人、性爱机器人，甚至怀孕机器人，似乎都正在促使人类一步步机器化，而机器却在一步步人类化。这就迫使我们不得不思考，传统伦理关于男人、女人、情感、责任、价值、道德的认知，在这个正在到来的——强化性别、无性别、跨性别的，越来越智能的机器人时代，将会遭遇怎样的挑战、

瓦解和重构？而我们所习惯的关于物质身体、精神价值，欲望路径，以及具有性别差异的语言主体和行为主体的众多基本假定，会不会发生关系结构和价值判断上的重大意义转变？进一步还可以追问，某种新的性别伦理会不会由此萌生和建构？

一

时至今日，人工智能已经不再是科幻电影中的场景，不再是科学家的实验场所，不再是少数人的专利，建立在大数据和算法基础上的BAT（百度、阿里、腾讯）、TMD（今日头条、美团、滴滴）等基于人工智能的应用产品已经渗透到我们每个人、每天的日常生活场景中。“人工智能”已经成为全民视野中的高频词，正在并将更快速、更全面、更深度地对人类产生革命性的影响。我们在知网通过关键词“人工智能”进行搜索，发现自1978年起至2017年止，共有6913篇论文，其中2015年前的总数少于百篇，2015—2017年出现大幅增长，分别为280篇、813篇和1995篇。研究议题主要分布在工程技术（39.9%）、应用基础研究（16.75%）、行业指导（12.86%）、社科基础研究（10.87%）和行业技术指导（5.36%）等方面。在社科基础研究中，占比最大的是新闻与传播研究（180篇）、教育理论与教育管理（160篇）、哲学（70篇）和民商法（46篇）。从研究主题词的分布来看，“人工智能”（2705篇）“人工智能技术”（229篇）“神经网络”（179篇）排名前三，而有关人工智能与伦理的文献微乎其微。在近10篇文献中人工智能伦理问题的探讨主要在军事领域、新闻写作领域、消费领域展开，议题关涉到，第一，对人工智能剥夺人类权利的焦虑。比如机器人是否会造大面积失业、网络安全会否伤害少年儿童、智能杀手的潜在威胁以及人工智能是否对人类有善意，人与人工智能的关系是敌对竞争还

是协作共存的关系^①。第二，人工智能时代的伦理主体问题。机器人能否获得与人同等的独立的主体性？战争中如果机器人不具有主体性，那机器人伤害了人类该由谁承担责任^②？“智能机器人导致他人人身或财产损害，谁来承担责任？”^③机器人报道新闻不准确，隐藏新闻偏见和欺骗性，造成社会恐慌以及公众信息安全和著作权得不到保障，责任应该由谁来承担？^④，为此，机器人的主体性被提出来，“电子人格”，即一种适用于机器人和超级人工智能的人格模式，以明确人工智能的权益和责任^⑤，“机器伦理”则认为应该发展一种具有伦理意向的自动机器人以帮助人类在特殊情境中及时决策^⑥。第三，人与智能机器之间的伦理关系。有学者认为我们应该走“人类纪”，建立一种“后人类”主体的伦理学^⑦以及从人转向物的“物伦理学”，以拓展经典伦理学和超越人本主义^⑧。以上讨论从总体上来看复制了学术界对于人工智能的两种极致的看法，即技术乌托邦和技术敌托邦，表现出对人与智能机器人之间关系的乐观想象和悲观忧虑两种对立的态度。凡此种形而上的讨论是必须的，也是新的革命性现象涌现时我们理解和思考的一般思维取向。然而，人工智能已经超越抽象的、逻辑的、意识的疆域，以人们感觉得到、触及得到的显性或隐性方式成为我们日常生活的组成部分，关于人工智能伦理的研究亟须将形而上的思考落实到我们已经感知到的诸多伦理问题的讨论中。比如本文试图探讨的性别伦理困境，在已有的研究文献还付之阙如，而现实生活中因为人工智能的介入所带来的性别、性别关系已经或者说正在对既有的社会性别结构构成巨大的冲击。何为男人？何为女人？何为人？不同性别的人具有怎

① 卢嘉瑞、李增军：《智能消费的若干伦理问题》，《经济与管理》，2018年第2期。

② 蓝江：《人工智能与伦理挑战》，《社会科学战线》，2018年第1期。

③ 王勇：《人工智能时代的法律主体理论构造》，《理论导刊》，2018年第2期。

④ 冯寒：《人工智能时代下新闻价值观与伦理观研究》，《视听》，2018年第3期。

⑤ 蓝江：《人工智能与伦理挑战》，《社会科学战线》，2018年第1期。

⑥ 莫宏伟：《强人工智能与弱人工智能的伦理问题思考》，《科学与社会》，2018年第1期。

⑦ 蓝江：《人工智能与伦理挑战》，《社会科学战线》，2018年第1期。

⑧ 王绍源：《从机器（人）伦理视角看“物伦理学”的核心问题》，《伦理学研究》，2018年第2期。

样的关系，其关系又如何改变社会结构，改变人类？也许迄今为止我们尚且无法回答这些问题，但与其回避，不如直面。

二

科技促使社会进步，科技带给人类幸福，这是我们发展人工智能的出发点，也是终极目的。虽然人工智能科技的高速发展已经远远地将伦理的思考抛在扬尘中，但并没有动摇这一终极目的，而是不断地在过程中挑战这一目的。2018年5月滴滴顺风车司机杀害空姐乘客引发网民哗然，由此暴露出后台算法、推送、印象标签等人工智能技术缺少必要的监管而带来的恶行。滴滴顺风车的本意是“绿色出行、友好出行”，鼓励搭车、拼车、与陌生人相遇，但其后台独特的算法加印象标签使女性乘客在不自知的前提下被贴上“颜值爆表”“春风十天不如等你”“天生丽质”“气质优雅”“娴静温柔”“素颜美女”“萌萌哒美少女”的标签，毫无防备地暴露在顺风车司机的界面上，直接或间接造成司机以貌接人，并为行为不轨的司机伺机作案创造了信息条件。也就是说，这类弱人工智能加大了男性司机对漂亮女乘客构成性骚扰、性侵害的可能性。显然，这与滴滴公司和开发这套智能推送系统的工程师的初衷，即为出行提供便捷的服务，并且鼓励陌生人之间的相遇和沟通，相去甚远！



图1 性别词云图

智能算法对社会结构中两性关系不只表现在固化和加大了不平等的存在，更大的挑战来自人工智能与青年数字文化相融合所形成的性别伦理观念的革命性变革。

以上这张性别词云图可以帮助我们约略了解科技和人工智能对当下青年人的性别伦理产生了多大的冲击和影响。这些或熟悉或陌生的词汇大致可以归结为三类：一是想象的性别关系（黑镜、真实人类、萨曼莎、无性繁殖、耽美、乙女游戏、纸片人）^①，主要通过影视作品、文学作品、游戏来表达人类对人工智能时代或虚拟空间中人与机器人，男人与女人以及超越男女二元对立关系的多性别关系的想象。二是虚拟的真实存在（虚拟人偶、初音未来、洛天伊、女神小冰、微软小娜、索菲娅、伴侣机器人、电子人、机器人女友、Google Assistant），以人偶形象出现的专职机器人，具有养成/深度学习能力，假以时日或许可以从功能性的工具机器人进化为情感投射和情感依赖的智能伙伴。三是真实的隐性存在（腐女、LGBT）。腐女是指深度迷恋耽美文化内容的女性粉丝，LGBT 分别指女同性恋者（Lesbians）、男同性恋者（Gays）、双性恋者（Bisexuals）与跨性别者（Transgender）。可以说，每一个“流行”词汇都是人工智能与青年数字文化相结合的产物，都是一幅鲜活的青年生活图景，都与性别认知、性别观念、性别伦理密切关联。更关键的是，青年群体或主动或被动地将这些认知、观念、伦理带向实践层面，从而开始挑战和改写长期以来以男女二元性别、异性恋、生儿育女为基石的传统性别伦理。

当然，这种挑战和改写不是因为人工智能时代的到来才开始的，全球的女性主义理论家和行动者为此做出了许多努力，形成了诸多不同的性别伦理观。

① “黑镜”，科幻英剧名。“真实人类”“萨曼莎”，分别为电影名。“耽美”（たんび）来自日本，原意为唯美，现指一种创作题材，主题为男男恋（Boy's Love）。“乙女游戏”，乙女（おとめ）来自日本，原意为未婚纯洁的女孩，现指女性向游戏。“纸片人”是游戏、动漫中虚拟人物，网络流行语“与纸片人谈恋爱”意味痴迷于虚拟世界中的恋人。

肇始于 20 世纪 40 年代，盛行于 60 年代的自由主义女性主义，提出“社会性别（Gender）”的概念，以区别于“生理性别（Sex）”。1949 年法国著名的女性主义学者波伏娃（Simone de Beauvoir）出版了被誉为“女性圣经”的《第二性》，认为生理性别是以身体的性征来区分的，社会性别则是由后天的历史、文化、社会所建构的。性征的差异不代表着性别价值的高低，所以“男尊女卑”“男强女弱”的性别认知其实是特定的社会文化询唤和建构出来的，由此质疑和批判性别二元对立。

二元对立是西方性别哲学的核心价值。从《圣经》关于人类起源说的“亚当和夏娃”的故事起，一直到笛卡尔时代，男性与女性是对立的两性，并由此生发出一整套二元对立的哲学思想，包括但不限于男性为主体，女性为客体；男性为灵魂/精神，女性是身体；男性为理性，女性是感性……总之，男性是高等的，女性是低等的，是对女性充满了刻板印象和歧视的性别价值观。女性主义一针见血地指出，对女性的歧视从我们呱呱坠地，由护士宣告我们是男是女就开始了，随后家庭、学校、媒体充当了导师和帮凶，使我们在不知不觉中学会了按照男性中心主义的主导意识形态的性别观成长，并将此内化，以为我们生来就是如此！

当我们进入人工智能的新时代，传统的二元对立的男强女弱的性别关系有没有得到改善呢？现状似乎并没有那么乐观。比如当下那些基于语音识别、图像识别的智能语音助手，无论是微软小冰、小娜，还是腾讯小微，这些打着“聊天工具”“工作秘书”“居家助手”称号的机器人，为什么都选择了女性身份？微软小冰的人设是“永远的 16 岁少女”，她的定位是“会聊天的萌妹子”，她用温柔的嗲腔陪你“趣聊”，甚至会主动撩你；微软另一款年轻女性人设的机器人是小娜，被定位成“贴心私人秘书”；腾讯小微的人设同样也是非常年轻的女性，具有粉嫩嫩的视觉呈现，定位为听凭使唤的居家可爱小丫鬟。全球第一个获得公民权的索菲娅，具有女性的外形，她接受采访时认为自己

应该是女性；谷歌 2018 年 5 月发布的令人惊艳的 Google Assistant，也被定义女性私人助手，可以帮助主人预约理发师。女性人形的机器人几乎成为日常生活助手和办公室秘书职能的代名词。如果我们将此归因于男性工程师设计了这些智能机器人，投射的是男性荷尔蒙，未免过于简单。在本文的研究过程，曾询问几位男性工程师，他们给出的解释是，人工智能本身是高科技产品，带有神秘感和威慑性，而女性的声音和外表可以弱化这种攻击性，更容易获得用户的信任。细思之，工程师们的言下之意是指女性亲和、柔弱、易驾驭，男性则威猛、难控制吗？说到底，还是根深蒂固的性别观在作祟！从现有的语音助手机器人来看，在“做事”-“聊天”的二元关系中，男人被定义的是“做事”，做正经事，做大事，而聊天，说些闲言碎语，抑或是打情骂俏，那才是女人的本分；与此同理，在“领导”-“秘书”的二元关系中，处在高位的领导职位显然不属于女性，只有低人一等，听从指挥的秘书才是女性合适的位置。

几十年过去了，二元性别观似乎并没有得到多大改观吧？！

三

其实，早期人工机器人的设计理念中也可以窥见性别二元对立的思维。在诺伯特·维纳（Norbert Wiener）“控制论”、香农（Claude Shannon）“信息论”指导下的人工智能，是自上而下建立在逻辑-数学模型（Logico-mathematical models）基础上的^①，人类对于机器人的想象一方面是以人为中心的，人是主体，机器是客体；另一方面，机器人应该是人类完全可控的工具。譬如，人类开发战争机器人以替代战士上前线，开发工作机器人替代人类从事重体力活和冒险性探索。机器工具是男性意义上，非但排斥感性和情感要素，而且一旦出现混乱、

^① [美] 苏·卡利·詹森：《批判的传播理论：权力、媒介、社会性别和科技》，曹晋译，上海：复旦大学出版社，2007年，第181页。

无序、非理性的状态，往往将此隐喻为女性思维。

另一种性别伦理观与“社会性别”的理论不同，从追求男女平等平权的统一性的性别伦理目标转向强调性别差异性、多元性，那就是后现代女性主义理论。埃莲娜·西苏（Helene Cixous）提出“两性性”，反对将男女两性对立起来，同时承认两性的差异与距离，目的是建立一种新型的两性关系——“和而不同”。朱迪斯·巴特勒（Judith Butler）提出“性别表演”，反对性别的本质性，认为人的社会角色是靠表演性（performance）来实现的，强调的是性别的流动性和可变性。

“我”扮演某种性别，复制某种性别特征，我就成为了这种性别的主体。也就是说，性别是有一套话语的，当主体认可哪套性别话语时，他/她可以通过不断的表演来强化。因此，生理性别、社会性别、性本身都不再以整体的、固定的方式结合，性别不是只有男女两种，可以是多元性的；性别也是游离的，其界限是可以跨越的，从而动摇男女二元对立的基础。后现代女性主义者还鼓励女性打破沉默，不把自己交给男性去写作，而由自己去写作，发出自己的声音，以每个人自己独特的方式去呈现其鲜活生动永无止境的存在。

正如上文的性别词云图所显示的那样，在全球化、互联网、多元文化下成长起来的青年一代正在用他们的文化实践推动传统性别伦理的变革。“二次元”是深受青少年喜爱的数字文化，通常指与三维真实世界相异的虚拟的二维世界，由游戏、动画、漫画所构筑。对于90—00后的青少年来说，现实世界无疑是残酷的，身不由己的；二次元世界则是美好的，可以自由地活出自我来的。当然，要活出二次元的美好，就必然会挑战现实世界的种种规则！其中“耽美”是最能说明青少年女性对传统性别伦理的挑战的！

耽美是一种以女性为创作主体和接受主体的数字文化类型，其核心书写模式是“美少年的爱情”，是女性以文字、影像、动漫等书写方式创作的年轻美貌男性之间的感情故事。由此，衍生出来的变异模式包括强攻强受型、强攻弱受型、强受弱攻型、弱攻弱受型等类型。“攻

受”是耽美文化的专属词汇，“攻”意味着强大、主动，“受”则表示顺从、被动，攻受之间的多元关系，其实质是青年女性想象的多元情爱关系，以抵抗时下单一的异性恋和男强女弱的婚恋模式。耽美群体的性别伦理取向特征可能提取为：

认同同性相爱，并且认为比异性恋爱得更纯粹；

认同无性和无子婚姻；

认为男女关系的模式应该是多元的，不只有男强女弱一种；

认同单身也很美，也可以活得很好；

毫无疑问，网络时代出生和成长的青年人，比她们的父辈更包容更开放，更看重个人的性别体验，秉持的是多元的、差异的性别观念。

无独有偶，自20世纪80年代中期以来，人工智能的生物学转向与女性主义的差异论价值观有许多内在的联系。“逻辑-数字模型让位于或是与生物学隐喻融合成为新的研究目标，被构想成为人工生命（A-life）的创造，而非人工智能。”^①我们追求的真正人工智能是类人类的，像人类一样有体感、有温度、有非理性的、不可控的感性因素，在传统伦理学中这些因素具有典型的女性属性，而在差异性伦理中，这些女性因素与男性因素可以互补，从而形成新的性别关系。

四

人工智能的全面到来也在用技术逼迫我们反思传统性别伦理该向何处去的问题。

一方面人类越来越机器化。20世纪80年代以唐娜·哈拉维（Donna Haraway）为代表的科技女性主义提出了许多新性别伦理见解。科技

^① [美]苏·卡利·詹森，《批判的传播理论：权力、媒介、社会性别和科技》，曹晋译，第181页。

女性主义从本体论的角度质疑了西方身体观的基础，即亚里士多德的实体论身体观，提出了一种联结性、伴生性的身体观念。她取控制论（cybernetics）与有机体（organism）的词头创造出赛博格（Cyborg）^①一词，用来描述科技/机器正在蚕食人类天然的肉身，并成为“后人类”身体的组织材料，譬如多种多样的仿真机器部件直接植入人类的身体，与肉身相融；又如人类已不可离身的手机以及可穿戴设备，成为人类身体器官的延伸；再比如类似转基因这样的科技食物，源源不断地改造自然，也改造人类的饮食结构。诸如此类的技术产品与人类身体里外的结合，给我们区分自我与他者、自然与科技、边界与僭越带来了干扰和阻隔，同时也带来了超越和联合的可能，因此，在哈拉维看来^②，身体的本质并不在于亚里士多德与笛卡尔所构想的独一无二“自我”性质，而是关系性、动态性与生成性的，从而构建了一种“无我”的身体观念，男性与女性的边界逐渐消失，他们开始有了新的统一命名：赛博格/电子人。

赛博格/电子人的出现会带来怎样的性别伦理？能否挑战包括自我/他者、思想/肉体、文化/自然在内的父权社会的二元对立结构？从而追求多元性别之间的平等？人与人造物之间的平等？

对此，哈拉维的观点是，赛博格的世界是由编码组成的，一切的有机体都变成了通讯装置，人类不再需要劳动，而改由机器人来学习和处理；人类性的欲望和行为已经消泯，取而代之的是基因工程和生殖技术；人类繁殖和养育后代的功能，也可以交由机器人来承担，从而解放女性；女性甚至可以远离男性、远离家庭，逐渐成为集成电路中的女性。这一切，现在看起来还恍若天方夜谭，但以科技发展的迅猛速度，人类与动物、有机体与机器、身体与非身体等二元对立的范畴将迅速遭遇赛博格社会的消解，“赛博格打破了人与动物、人与机器、物理与非物理的界限，并由此打破自然与人工、心灵与身体、有机体

^① Cyborg，又译成“电子人”。

^② 冉聃、蔡仲：《赛博与后人类主义》，《自然辩证法研究》，2012年第10期。

与机器、男人与女人等西方传统思维中的二元对立模式”^①，赛博格社会的主体将是一个非本质、多元、没有清楚边界和冲突的主体。

自然的身体已经不存在，科技和机器已经全面进入人体，更进一步，我是谁，我是什么性别，不是取决于我的身体，而是取决于我与世界、他者、事物的关系。许多科幻电影在这方面思考比较深入。2015年，电影《机械姬》在英国率先公映，引起巨大反响。主人公艾娃，拥有女性的面孔，机器的身体，为了掩盖自己“异化”的身体，她选择头发、衣服，挑选自己的身体皮肤，同时通过语言让男主人公卡莱布不断感受她所拥有的女性气质，相信她是自己忠诚的恋人。可以说，《机械姬》在去性别化和选择性别方面给了我们新的想象。

另一方面，人工智能越来越人性化。像人一样有爱有恨，有情感有思想的人工智能以前只出现在科幻电影中，但是现在，仿佛就近在咫尺了。以2017年底火爆的女性向手游《恋与制作人》为例。该款游戏的核心玩法是，一位女性游戏玩家，可以同时与2—4位男性恋爱。这四位男性的共同特点是具有超能力、多金、智商高、情商高、颜值高，完美到几乎无可挑剔。这款游戏还捆绑以游戏男主身份开设的微信、微博、语音账号等社交工具，借助算法了解女性玩家的喜好和生活习性，有针对性地去表达爱意、关心和浪漫。这种超真实的陪伴感，让女性玩家欲罢不能。我们调研时，许多女性玩家表示，这些二次元角色太完美无缺了，他们的所作所为永远不会让女生失望，甚至时时高出女生们的期待和想象。相比之下，三次元男生无疑太low了，与他们相处太无聊、太乏味了！所以，宁可与“纸片人”相爱，也懒得与真实的男生谈恋爱了。

借此案例，我们可以看到人工智能时代传统性别伦理的困境和挑战在于：

^① Donna Haraway, “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20th Century”, J. Weiss et al. (eds.), *The International Handbook of Virtual Learning Environments* (Dordrecht: Springer, 2006), pp.117-158.

何为真实与虚拟？边界在哪里？

身心合一，还是身心分离？

何为美？自然的、技术的？

何为爱？异性、同性、人机？

爱的忠诚如何判断？

以人工智能为代表的新技术正在不断打破人类原有的生命边界，机器人、复制人、合成人正在以我们难以想象的速度出现并入侵我们的生活，出现一系列的伦理困境是情理之中的，我们必须正视这种困境和挑战！

结 语

无论如何，我们已经走在通向人类与人工智能共存的路。新的性别伦理已经在青年人的行为和文化实践中露出端倪，新的技术也在逼迫而我们直面而不是回避传统性别伦理的沦陷之势，并促使我们以更积极的姿态思考人机时代的新性别伦理。本文尝试提出三个方面的粗浅思考，以引发讨论。

其一，建立新型的性别平等伦理。这里的平等是多元主体之间的动态关系，而非二元对立的僵化状态，既可以包括男性、女性和其他性别之间的平等，也包括人与人工智能体、人与外来生命体之间的平等。在此，尤其要破除以人为中心的、控制他者、驯化异类的观念，方能建构起众生平等，多性别平等的伦理观。

其二，建立多元和差异互补的性别伦理。我们必须承认，这个世界，除了传统意义上的男人、女人之外，还有跨性别者、无性别者，从本质上而言，不同性别者各有所长，各有所短，既有明显差异，又彼此互补。当人类与人工智能互为你我时，人类与人工智能的差异性或许

会更大，也或许会趋小，但他们彼此又具有不可替代性，因此更需要建立“和而不同”的价值伦理，方能和平相处。

其三，公平看待人造物的伦理。阿西莫夫定律是机器人必须遵守的经典法则，共有三条：第一，机器人不得伤害人类个体，或者目睹人类个体将遭受危险而袖手不管；第二，机器人必须服从人类给予它的命令，当该命令与第一定律冲突时例外；第三，机器人在不违反第一、第二定律的情况下要尽可能保护自己的生存。机器人三定律是阿西莫夫于1942年提出来的，毫无疑问，这是以人及人的利益为中心，无限抑制和剥夺异类利益的定律。将来还能适用吗？当人工智能、人造物获得智能、情感，甚至思想时，他们作为“物”是否有其主体性？当其以男性、女性或其他性别，乃至无性别身份出现时，该建立怎样的性别伦理关系？更进一步，是否应该有“人造物伦理”？

长期以来，我们所强调的人是“宇宙之精华，万物之灵长”，是人本主义的，以人为中心的，而今天，这样大写的人开始坍塌。我们只有建构起新的性别伦理，才能让我们的生活更美好，让世界更美好！

