第五篇　伦理篇：人类价值与人机关系

随着第三次浪潮开启的人工智能时代的来临，人工智能伦理再次成为各界热议和研究的核心议题之一。联合国、欧盟、美国、英国等都格外重视这个问题，纷纷出台研究报告、指南、法律政策等多种措施，推进对人工智能伦理问题的认知和解决；电气与电子工程师学会（IEEE）、国际互联网协会、阿西洛马会议以及业内的谷歌、IBM等亦开始以伦理标准、人工智能原则以及人工智能伦理审查委员会等行业自律的形式积极应对人工智能伦理问题。

可以预见，随着机器智能的崛起，人工智能、机器人等开始从事越来越多的道德决策，以及未来强人工智能和超人工智能将带来人机区分的难题，加强人工智能伦理研究将显得尤为重要，尤其是为了保存人类的价值以及实现人机共存共荣的美好愿景。

第二十四章　道德机器

我们最好相当确信，植入机器中的目的就是我们真正想要的目的。

——诺伯特·维纳

2016年可谓是人工智能史上异常耀眼的一年。这一年，谷歌公司的DeepMind团队开发的围棋机器人程序AlphaGo首次击败顶尖人类棋手，深度学习、强化学习等人工智能技术功不可没。次年，AlphaGo在多个场合横扫几乎所有顶尖人类棋手，人类统治围棋的时代彻底告终。另一机器人程序Libratus在德州扑克比赛中击败顶级人类玩家，这是机器人首次在不完全信息博弈中战胜人类。这些事件标志着机器智能的崛起，人类社会正在逐步进入智能机器的时代，机器辅助甚至取代人类进行各种决策将越来越常见和普遍。

此次人工智能浪潮的标志便是深度学习，是能够自我学习、自我编程的学习算法，可以用来解决更复杂的认知任务，而这些任务在此前完全专属于人类或者人类专家，比如开车、识别人脸、提供法律咨询服务，等等。深度学习、强化学习等机器学习技术，结合大数据、云计算、物联网以及其他软硬件技术，使得机器智能取得重大突破。在此背景下，有关道德机器的呼声再起。

智能机器加速到来

人工智能技术助力智能机器加速到来，机器逐步从被动工具向能动者转变。“计算机仅能执行强制的指令——对其编程不是为了使其能够作出判断。”纽约一家法院曾经如是说。这或许可以代表公众对计算机和机器人的固有看法。但是，人工智能技术的进步，正使这一观点变得陈腐，甚至可能成为一个偏见。因为机器正从被动工具向能动者转变，可以像人一样具有感知、认知、规划、决策、执行等能力。承担义务，并对其造成的损害承担责任？这些都是欧盟未来在对机器人立法时需要重点考虑的问题。

2010年以来，受到大数据、持续改进的机器学习、更强大的计算机、物理环境的IT化（物联网）等多个相互加强的因素推动，人工智能技术在ICT领域快速发展，不断被应用到自动驾驶汽车、医疗机器人、护理机器人、工业和服务机器人以及互联网服务等越来越多的领域和场景。国外一些保险和金融公司以及律师事务所甚至开始用具有认知能力的人工智能系统替换人类雇员。从国际象棋、智力竞赛（比如*Jeopardy*），到围棋、德州扑克，再到医疗诊断、图像和语音识别，人工智能系统在越来越多的领域开始达到甚至超过人类的认知水平，让其辅助甚至代替人类进行决策，不再是空中楼阁。现在有理由预见，在不远的将来，交通运输、医疗、看护、工业和服务业等诸多领域的各式各样的智能机器（Intelligent Machine）或者智能机器人（Smart Robot）将成为人类社会中司空见惯的事物。

以自动驾驶汽车为例，其区别于传统的机器的最大特征在于具有高度的甚至完全的自主性。无论采用何种机器学习方法，当前主流的深度学习算法都不是一步一步地对计算机

编程，而是允许计算机从数据（往往是大量数据）中学习，不需要程序员作出新的分步指令。因此，在机器学习中，是学习算法（Learning Algorithm）创建了规则，而非程序员。其基本过程是给学习算法提供训练数据，然后，学习算法基于从数据中得到的推论生成一组新的规则，称之为机器学习模型。这意味着计算机可被用于无法进行手动编程的复杂认知任务，比如图像识别、将图片翻译成语音、汽车驾驶等。就自动驾驶汽车而言，其利用一系列雷达和激光传感器、摄像头、全球定位装置以及很多复杂的分析性程序和算法等，像人类一样驾驶汽车，而且做得更好。自动驾驶汽车“观察”路况，持续注意其他汽车、行人、障碍物、绕行道等，考虑交通流量、天气以及影响汽车驾驶安全的其他所有因素并不断调整车速和路线。而且自动驾驶汽车被编程来避免与行人、其他车辆或者障碍物发生碰撞。所有这一切都是机器学习的结果。因此可以说，在每一个现实情境中，都是自动驾驶汽车自身在独立判断和决策，虽然是程序员设定了学习规则。

更进一步，智能机器可能“打破”预先设定的规则，大大超出其设计者的预期。人们一直担心，赋予机器自主“思考”的能力可能导致其有能力违反被设定的“规则”，以人们意想不到的方式行为。这不纯粹是想象，已经有证据表明高度“智能”的自主机器可以学习“打破”规则以保护其自身的生存。自动驾驶汽车脱离制造商控制，继而进入流通领域之后的学习和经历同样影响其行为和决策。新的数据输入可能使自动驾驶汽车进行调整和适应，导致其行为和决策超出预先设置的规则，这在理论上并非不可能。

这些现象无不表明，计算机、机器人、机器等正在脱离人类直接控制，独立自主地运作，虽然它们依然需要人类进行启动，并由人类对其进行间接控制。但就本质而言，自动驾驶汽车、智能机器人、各种虚拟代理软件等都已经不再是人类手中的被动工具，而成为了人类的代理者，具有自主性和能动性。这对伦理和道德提出重大挑战，之前针对人类和人类社会的伦理规范现在需要延伸到智能机器，而这可能需要新的伦理范式。

道德代码的必要性

未来的自主智能机器将有能力完全自主行为，不再是为人类所使用的被动工具。虽然人类设计、制造并部署了它们，但它们的行为却不受人类的直接指令约束，而是基于对其所获取的信息的分析和判断，而且，它们在不同情境中的反应和决策可能不是其创造者可以预料到或者事先控制的。完全的自主性意味着新的机器范式：不需要人类介入或者干预的“感知-思考-行动”。这一转变对人工智能、机器人等提出了新的伦理要求，呼吁针对机器的新的伦理范式。

当决策者是人类自身，而机器仅仅是人类决策者手中的工具时，人类需要为其使用机器的行为负责，具有善意、合理、正当使用机器的法律和伦理义务，在道义上不得拿机器这一工具来从事不当行为。此外，除了善意、正当使用工具的义务，当人类决策者借助工具来从事不当或者违法行为时，人类社会一方面可以在道德和舆论层面对其进行谴责，一方面可以借助法律这一工具对违法者进行惩罚。然而，既有的针对人类决策者的法律和伦理路径并不适用于非人类意义上的智能机器。但是，由于智能机器自身在替代人类从事之前只能由人类做出的决策行为，因此在设计智能机器时，人们需要对智能机器这一能动者提出类似的法律、伦理等道义要求，确保智能机器做出的决策可以像人类一样，也是合伦理、合法律的，并且具有相应的外在约束和制裁机制。

更进一步，智能机器决策中的一些问题也彰显了机器伦理的重要性，需要让高度自主的智能机器成为一个像人类一样的道德体，即道德机器（Moral Machine）。其中一个问题是，由于深度学习算法是一个“黑箱”，人工智能系统如何决策往往并不为人所知，其中可能潜藏着歧视、偏见、不公平等问题。人工智能决策中越来越突出的歧视和不公正问题使得人工智能伦理显得尤为重要。尤其是人工智能决策已经在诸如开车、贷款、保险、雇佣、犯罪侦查、司法审判、人脸识别、金融等诸多领域具有广泛应用，而这些决策活动影响的是用户和人们的切身利益，确保智能机器的决策是合情合理合法的就至关重要，因为维护每个人的自由、尊严、安全和权利，是人类社会的终极追求。

此外，战争中的人工智能应格外受到伦理规范约束。目前许多国家都在积极研发军用机器人，而军用机器人的一个重要发展趋势就是自主性在不断提高。比如，美国海军研发的X-47B无人机可以实现自主飞行与降落。韩国、以色列等国已经开发出了放哨机器人，它们拥有自动模式，可以自行决定是否开火。显然，如果对军用机器人不进行某种方式的控制的话，它们很可能对人类没有同情心，对目标不会手下留情，一旦启动就可能成为真正的冷血“杀人机器”。为了降低军用自主机器人可能导致的危害，需要让它们遵守人类公认的道德规范，比如不伤害非战斗人员、区分军用与民用设施等。虽然现有技术要实现这样的目标还存在一定的困难，但技术上有困难并不意味着否定其必要性与可能性。[1]

道德机器的实现

机器人、智能机器等人工智能系统需要遵守人类社会的道德、法律等规范并受其约束，但如何实现这一目的，即设计出道德机器，将人类社会的法律、伦理等规范和价值嵌入人工智能系统，是一个很大的挑战。首先，人们需要发问，法律、道德等要求和规范可以被转化成计算机代码吗？也即道德、伦理的计算机代码。其次，如果可以，需要嵌入的规范和价值是什么？以及应以怎样的方式将这些法律和道德的要求嵌入人工智能系统？最后，如何确保嵌入人工智能系统的规范和价值符合人类的利益并与时俱进？解决这三个问题，基本就可以确保实现机器伦理，让人工智能系统成为像人类一样善意、正当、合法行为的能动者。

为了解决伦理嵌入的问题，2016年底，IEEE启动了人工智能伦理工程，发布了《合伦理设计：利用人工智能和自主系统（AI/AS）最大化人类福祉的愿景》，从可操作标准的层面为伦理嵌入提供指引，值得探讨和借鉴。IEEE将人工智能伦理的实现分为三个步骤。

第一，识别特定社群的规范和价值。首先，应当明确需要嵌入AI系统的规范和价值是什么。法律规范一般是成文的、形式化的，容易得到确认。但社会和道德规范比较难确认，它们体现在行为、语言、习俗、文化符号、手工艺品等之中。更进一步，规范和价值不是普世的，需要嵌入AI的价值应当是特定社会或团体中针对特定任务的一套规范。其次，道德过载（Moral Overload）问题。AI系统一般受到多种规范和价值约束，诸如法律要求、金钱利益、社会和道德价值等，它们彼此之间可能发生冲突。在这些情况下，哪些价值应当被置于最优先的地位？因此，应优先考虑广大利益相关方群体共同分享的价值体系；在AI研发阶段确定价值位阶时，需要有清晰、明确的正当理由；在不同情境下或随着时间的推移，价值位阶可能发生变化，技术应当反映这一变化。最后，数据或算法歧视问题。AI系统可能有意或无意地造成对特定使用者的歧视。一方面，要承认AI系统很容

易具有内在歧视，意识到这些歧视的潜在来源，并采取更加包容的设计原则；强烈鼓励在整个工程阶段，从设计到执行到测试再到市场推广，尽可能具有广泛的包容性，包容所有预期的利益相关方。另一方面，在解决价值冲突时保持透明性，尤其需要考虑弱势的、易被忽视的群体（儿童、老年人、罪犯、少数民族、贫困人群、残障人群等）的利益；在设计过程中，采取跨学科的路径，让相关专家或顾问团体参与其中。

第二，将发现并确定的规范和价值嵌入人工智能系统。在规范体系得到确认之后，如何将其内置到计算机结构中，是一个问题。虽然相关研究一直在持续，这些研究领域包括机器道德（Machine Morality）、机器伦理学（Machine Ethics）、道德机器（Moral Machine）、价值一致论（Value Alignment）、人工道德（Artificial Morality）、安全AI、友好AI等，但开发能够意识到并理解人类规范和价值的计算机系统，并让其在做决策时考虑这些问题，一直困扰着人们。

当前主要存在两种路径：自上而下的路径和自下而上的路径。瓦拉赫与艾伦把“自上而下”方法描述为“利用特定的伦理理论进行分析，指导实现该理论的运算法则（Algorithms）和子系统（Sub-System）的计算需要”的方法：把“自下而上”方法称作拓展式方法，其指出该方法“重点在于为主体探索行动和学习方面营造一个环境，鼓励其实施道德可嘉型行为”。他们宣称，自下而上型方法的优点在于其能够“从不同的社会机制中动态地进行集成输入”，能够为完善其整体发展提供技巧和标准，但这一方法可能存在“很难适应和发展”的弊端。[2]目前还尚未明确如何将这些规范嵌入到计算机架构中，这一领域的研究需要加强。

但是，这里可能依然存在一个难以达成共识，但却需要事先解决的伦理困境。以自动驾驶汽车为例，我们可以假设一个类似“电车困境”的伦理问题。如果一辆自动驾驶汽车在刹车失灵或者来不及刹车的情况下，正好道路前方有五人闯红灯，而车上有两个乘客，此时，如果继续前行则会撞死不遵守交通规则的五人，而如果转向则会碰到路障，导致车上两人丧生。在此情形下，人们应当期待该汽车如何选择呢？由于人类自身的伦理价值有时候是似是而非，或者相互冲突的，自动驾驶汽车此时可能难以做出公认为正当的选择。

比如，按照功利主义，本着最大化最大多数人的利益和福利的目的，该车应当牺牲车上两人，而拯救闯红灯的五人。但是，按照绝对主义的道德要求，违背一个人的自由意志而伤害一个人的行为是不被允许的，不能为了拯救多数人，而违背其自由意愿伤害少数人，在这个情境下，就是使车上乘客丧生。解决这样的问题，对于自动驾驶汽车等人工智能系统的发展和商业化应用是非常重要的，所以全球各国都在积极关注和应对。[3]

第三，评估嵌入人工智能系统的规范和价值是否和人类的相符。因此，需要对嵌入AI系统的规范和价值进行评估，以确定其是否和现实中的规范体系相一致，而这需要评估标准。评估标准包括机器规范和人类规范的兼容性、AI经过批准、AI信任等。

在人类和AI之间建立信任涉及两个层面。一方面，就使用者而言，AI系统的透明性和可验证性对于建立信任是必要的；当然，信任是人类-机器交互中的一个动态变量，可能随着时间推移而发生变化。另一方面，就第三方评估而言，其一，为了促进监管者、调查者等第三方对系统整体的评估，设计者、开发者应当日常记录对系统做出的改变，高度可追溯的系统应具有一个类似飞机上的黑匣子的模型，记录并帮助诊断系统的所有改变和行为；其二，监管者连同使用者、开发者、设计者可以一起界定最小程度的价值一致性和相符性标准，以及评估AI可信赖性的标准。

人工智能伦理评估中更为重要的一个问题其实是价值对接。现在的很多机器人都是单一目的的，扫地机器人就会一心一意地扫地，服务机器人就会一心一意给你煮咖啡，诸如此类。但机器人的行为真的是我们人类想要的吗？这就产生了价值对接问题。可以举一个神话故事，迈达斯国王想要点石成金的技术，结果当他拥有这个法宝的时候，他碰到的所有东西包括食物都会变成金子，最后却被活活饿死。为什么呢？因为这个法宝并没有理解迈达斯国王的真正意图，那么机器人会不会给我们人类带来类似的情况呢？这个问题值得深思。

因为家庭服务机器人可能为了给你的孩子做饭，而杀死你家的宠物狗。更极端地，一个消除人类痛苦的机器人可能发现人类在即使非常幸福的环境中，也可能找到使自己痛苦的方式，最终这个机器人可能合理地认为，消除人类痛苦的方式就是清除人类，这一假设在医疗机器人、养老机器人等方面具有现实的影响。所以有人提出兼容人类的AI，包括三项原则：一是利他主义，即机器人的唯一目标是最大化人类价值的实现；二是不确定性，即机器人一开始不确定人类价值是什么；三是考虑人类，即人类行为提供了关于人类价值的信息，从而帮助机器人确定什么是人类所希望的价值。解决价值对接问题，需要更多跨学科的对话和交流机制。

人工智能伦理和道德机器的实现需要综合的治理模式

如前所述，有关道德机器的论证主要着眼于两个方面：一是关于道德和伦理的可操作标准；二是伦理工程的方法论。正是由于这两个问题的存在，人工智能伦理才成为一个跨学科的问题，需要跨学科的路径和方法，单靠人文学者或者技术人员是无法完成的。因为，跨学科的参与、对话和交流在未来应对人工智能伦理问题时，是极为必要的。

此外，正如人类通过学习、社会交往等习得道德、法律、伦理等规范和价值，并予以自我遵守一样，机器伦理也希望达成同样的效果。通过伦理标准的设定、执行、检测检验等，旨在希望以事前的方式让智能机器的自主决策行为尊重人类社会的各种规范和价值，并最大化人类整体的利益。

考虑到对于人类的行为，仅有人类的道德、法律自律是远远不够的，还需要一套外在的监督和制裁机制，因此将伦理嵌入人工智能系统这样一种自律的行为，也是远远不够的，还需要政府监管机构、社会公众等的共同参与，以事中或者事后的方式对人工智能系统的行为进行监督、审查和反馈，共同实现人工智能伦理，确保社会公平正义。因此，人工智能伦理的实现，是一项全方位的治理工程，需要AI研发人员、企业、政府、社会各界以及用户的共同参与，发挥各自不同的作用和角色，确保人工智能系统以尊重、维持人类社会既有伦理、法律等规范和价值的方式运作，带来最大化效益和好处的同时，也能够维护整个社会以及每一个个体的自由和尊严。

---

[1] 杜严勇.现代军用机器人的伦理困境.伦理学研究，2014（5）：98-99.

[2] 王绍源.论瓦拉赫与艾伦的AMAs的伦理设计思想:兼评《机器伦理：教导机器人区分善恶》.洛阳师范学院学报，2014，33（1）：32.

[3] 王东浩.人工智能体引发的道德冲突和困境初探.伦理学研究，2014（2）：70.

第二十五章　人工智能23条“军规”

计算机科学与人工智能之父艾伦·图灵曾说过，“即使我们可以使机器屈服于人类，比如，可以在关键时刻关掉电源，然而作为一个物种，我们也应当感到极大的敬畏。”图灵说下这段话的时间是1951年，彼时，人工智能的概念尚未诞生。但是，学术领袖对机器具有超越人类的智能，并可能威胁人类的担忧就已然存在，而且在人工智能技术及其应用突飞猛进的今天，依然具有重要的警示意义。伴随着人工智能可能控制、毁灭人类的担忧不断发酵，政府、业界和企业对人工智能“军规”“紧箍咒”的探索开始紧锣密鼓地进行，目的在于让人工智能造福于人类的同时，也是安全、可靠、可控的，不会威胁到我们这个物种的生存。

对机器失控的担忧由来已久

1956年夏季，杰出的计算机科学家们在美国东部城市达特茅斯召开会议，首次提出了“人工智能”的概念。在这次会议上，首次决定将像人类一样思考的机器称为“人工智能”。此后，人工智能就一直萦绕在人们的耳畔，经历了若干次的高潮与低谷。如今，第三次人工智能浪潮已经到来，其发展速度将会大大加快。在这一过程中，无论是《终结者》《黑客帝国》等科幻文学令人惊悚的叙述，还是霍金等科学家振聋发聩的警告，抑或是马斯克等业界领袖的担忧，都无不透露出人们对未来通用型人工智能和超级人工智能的担忧。

可以预见，人工智能正在从弱人工智能向通用型人工智能和超级人工智能方向发展。只要技术一直发展下去，人类终有一天会造出通用型人工智能，进入数学家I.J.Good提出的“智能大爆炸”或者“技术奇点”阶段，到那时，通用型人工智能有能力循环性地自我提高，导致超级人工智能的出现，而且上限是未知的。被比尔·盖茨誉为“预测人工智能最厉害的人”的库兹韦尔预言，2019年机器人智能将能够与人类匹敌；2030年人类将与人工智能结合变身“混血儿”，计算机将进入身体和大脑，与云端相连，这些云端计算机将会增强我们现有的智能；到2045年，人与机器将会深度融合，人工智能将会超过人类本身，并开启一个新的文明时代。

未来的通用型人工智能和超级人工智能一旦不能有效受控于人类，就可能成为人类整体生存安全的最大威胁；与核弹等原子核技术相比，这种威胁只会有过之而无不及，因此需要人类提前防范。虽然正如美国白宫人工智能报告《为人工智能的未来做好准备》所言，当前处在弱人工智能阶段，普遍人工智能在未来的几十年都不会实现，但人工智能领域很多研究人员都认为，只要技术持续发展下去，通用型人工智能以及之后的超级人工智能就必然会出现，主要的分歧点在于通用型人工智能和超级人工智能何时会出现。2016年以来，诸如霍金、伊隆·马斯克、埃里克·施密特等知名人士都对人工智能的发展表达了担忧，甚至认为人工智能的发展将开启人类毁灭之门。

霍金在演讲时认为，“生物大脑与电脑所能达到的成就并没有本质的差异。因此，从

理论上讲，电脑可以模拟人类智能，甚至可以超越人类。”伊隆·马斯克警告道，对于人工智能，如果发展不当，可能就是在“召唤恶魔”。人们担忧，随着普遍人工智能的发展，人类将迎来“智能大爆炸”或者“奇点”；届时，机器的智慧将提高到人类望尘莫及的水平。当机器的智慧反超人类，超级智能机器出现之时，人类将可能无法理解并控制自己的造物，机器可能反客为主，这对人类而言是致命的、灾难性的。这是一个值得深思的问题，当然也需要提前研究，采取防范措施，确保人工智能朝着有益、安全、可控的方向发展，而这就需要提出恰如其分的人工智能“军规”，给人工智能的发展套上“紧箍咒”，最大化人类的利益。

阿西莫夫机器人三定律靠谱吗？

最早提出应对人工智能安全、伦理等问题的非阿西莫夫莫属。阿西莫夫在多部科幻小说中，经常提到机器人的工程安全防护和伦理道德标准。在1942年问世的科幻小说《环舞》中，他提出了机器人三定律，以期对机器人进行伦理规制。

第一定律，机器人不能伤害人类，不能袖手旁观坐视人类受到伤害。第二定律，机器人应服从人类指令，除非该指令与第一定律相悖。第三定律，在不违背第一和第二定律的情况下，机器人应保护自身的存在。后来阿西莫夫对机器人三定律进行了补充，提出了第零定律，约定机器人必须保护人类的整体利益不受伤害。人们评价第零定律时认为，“人类整体利益这个混沌的概念，连人类自己都搞不明白，更不用说那些用0和1思考问题的机器人了。”

但人们始终质疑，机器人三定律能否真正解决机器人的安全、伦理等问题。正如阿西莫夫许多小说里显示的，机器人三定律的缺陷、漏洞和模糊之处将不可避免地导致一些奇怪的机器人行为。比如，在影片《我，机器人》中，VIKI机器人为了人类物种的延续，阻止人类之间的战争，最后决定限制人类的自由，就让人很难接受。但从三定律的内容来看，其并没有违反三定律，因为三定律并没有关于人权方面的定义，仅仅是保证人类生命的安全。所以影片中机器人限制人类自由来保护人类的行为完全是遵守定律的行为。

阿西莫夫的机器人系列科幻小说中还有很多机器人对于三定律之间发生矛盾和冲突的场景描述，人们可以看到机器人三定律对于构建机器人安全和伦理的缺陷和不足。机器人三定律存在的矛盾从日常生活中也能发现，比如警察和歹徒在枪战，按照第一定律，机器人不能袖手旁观坐视人类受到伤害，那么机器人必定需要帮助两方确保其不受到伤害，但这样的情形是人类希望看到的吗？诸如此类的很多场景，我们都会看到机器人三定律所存在的缺陷。

对于机器人三定律的意义，人工智能学家路易·海尔姆和本·格策尔也发表了一些看法。海尔姆认为超级人工智能必定会到来，而构建机器人伦理是人类面临的一大问题。他认为根据机器伦理学的共识，机器人三定律无法成为机器人伦理的合适基础，无论是AI安全研究者还是机器伦理学家都没有真正将它作为指导方案。原因是这套伦理学属于“义务伦理学”范畴，按照义务伦理学，行为合不合道德，只决定于行为本身是否符合几项事先确定的规范，和行为的结果、动机等毫无关系，这就使得面对复杂的情况时机器人无法作出判断或者实现符合人类预期的目的。格策尔也认为用三定律来规范道德伦理必定是行不通的，而且三定律在现实中完全无法运作，因为其中的术语太模糊，很多时候需要主观

解释。海尔姆认为机器伦理路线应该是更合作性、更自我一致的，而且更多地使用间接规范，这样就算系统一开始误解了或者编错了伦理规范，也能恢复过来，抵达一套合理的伦理准则。

对新一轮人工智能“军规”的探索

阿西莫夫的机器人三定律并没有给机器人的安全可控和伦理问题提供清晰的指引。由于之前的两次人工智能浪潮并未引起太多的关注，所以未能引起政府以及社会各界的广泛关注和担忧。然而这一次大为不同，人工智能将超越人类的呼声和担忧此起彼伏，各国政府、业界以及企业等开始积极关注、推进人工智能的安全和伦理，开始了对人工智能“军规”的新一轮探索。

第一，各国政府密切关注人工智能安全，出台安全和伦理举措。

人工智能不仅是以互联网为首的产业界竞相追逐的对象，而且是世界范围内的公共政策热议的焦点，各国政府及各组织纷纷开始了对人工智能的立法进程，目的之一便是加强人工智能安全。2016年8月，联合国世界科学知识与科技伦理委员会发布《关于机器人伦理的初步草案报告》，认为机器人不仅需要尊重人类社会的伦理规范，而且需要将特定伦理准则编写进机器人中。此外，英国政府2016年9月发布的《机器人技术和人工智能》呼吁加强AI伦理研究，最大化AI的益处，并寻求最小化其潜在威胁的方法。

欧盟也进行相关立法，为人工智能研发和审查人员制定伦理守则，确保在整个研发和审查环节中考虑人类价值，使其研发的机器人符合人类利益。2016年5月，法律事务委员会发布《就机器人民事法律规则向欧盟委员会提出立法建议的报告草案》；同年10月，发布研究成果《欧盟机器人民事法律规则》。在这些报告和研究的基础上，2017年2月16日，欧洲议会以396票赞成、123票反对、85票弃权通过一份决议，提出了一些具体的立法建议，要求欧盟委员会就机器人和人工智能提出立法提案（在欧盟只有欧盟委员会有权提出立法提案）。在其中，欧盟针对AI科研人员和研究伦理委员会（REC）提出了一系列需要遵守的伦理准则，即人工智能伦理准则（“机器人宪章”），诸如人类利益、不作恶、正义、基本权利、警惕性、包容性、可责性、安全性、可逆性、隐私等。此外，在安全方面，欧盟提出了一些基本原则，比如，因为机器人未来可能具有意识，因此阿西莫夫的机器人三定律必须传递给设计者、制造商和机器人操作者，因为这些定律不能转化为计算机代码。

在英国，2016年4月英国标准组织（BSI）发布机器人伦理标准《机器人和机器系统的伦理设计和应用指南》（BS 8611 Ethics Design and Application Robots），为识别潜在伦理危害提供指南，为机器人设计和应用提供指南，完善不同类型的机器人的安全要求，其代表了“把伦理价值观嵌入机器人和AI领域的第一步”。指南首先指出：“机器人的主要设计用途不能是杀人或伤害人类；应该由人类对事情负责，而不是机器人；对于任何一个机器人的行为都应该有找到背后负责人的可能。”指南建议机器人设计者要以透明性为导向，虽然这一点在实际设计中有困难。指南还提到机器人歧视等社会问题的出现，警告小心机器人缺乏对文化多样性和多元化的尊重。

第二，行业签署阿西洛马人工智能原则。

2017年1月，在加利福尼亚举办的阿西洛马AI会议上，特斯拉CEO埃隆·马斯克、DeepMind创始人戴米斯·哈萨比斯以及近千名人工智能和机器人领域的专家，联合签署了阿西洛马人工智能23条原则，呼吁全世界在发展人工智能的同时严格遵守这些原则，共同保障人类未来的利益和安全。霍金和马斯克公开声明支持这一系列原则，以确保拥有自主意识的机器保持安全，并以人类的最佳利益行事。

阿西洛马AI原则分为三大类23条（见图5-1）。第一类为科研问题，共5条，包括研究目标、研究资金、科学-政策连接、研究文化以及避免竞赛。主要内容包括：人工智能的研究目标不能不受约束，必须发展有益的人工智能；法律应该跟上AI的步伐，应该考虑人工智能“价值观”问题；人工智能投资应该附带一部分专项研究基金，确保人工智能得到有益的使用，以解决计算机科学、经济、法律、伦理道德和社会研究方面的棘手问题。此外，应该努力使研究人员和法律、政策制定者合作，并且应该在AI的研究人员和开发人员之间形成合作，培养整体的信任和尊重文化。

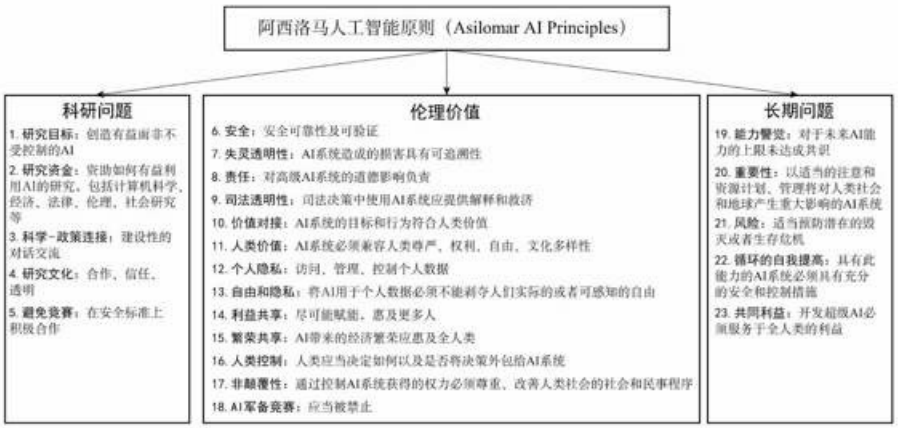

图5-1　阿西洛马人工智能原则

第二类为伦理价值，共13条，包括AI开发中的安全、透明度、责任、价值观等。主要内容包括：应该以安全和透明的方式研究AI；如果人工智能系统造成了损害，造成损害的原因要能被确定；在司法决策系统中使用任何形式的自动化系统，都应该提供令人满意的解释，而且需要由有能力的人类监管机构审核；对于先进的人工智能系统在使用、滥用和应用过程中蕴含的道德意义，设计者和开发者都是利益相关者，他们有责任也有机会塑造由此产生的影响；人工智能系统的设计和运行都必须与人类的尊严、权利、自由以及文化多样性的理想相一致。

第三类为长期问题，共5条，旨在应对AI造成的灾难性风险。主要内容包括：必须针对人工智能风险所引致的预期影响制定相应的规划和缓解措施；对于能够通过自我完善或自我复制的方式，快速提升质量或增加数量的人工智能系统，必须辅之以严格的安全和控制措施；超级人工智能只能服务于普世价值，应该考虑全人类的利益，而不是一个国家或一个组织的利益。

总而言之，阿西洛马人工智能原则对人工智能长期安全的担忧可以看作是对过去60多年公共话语的一个总结，也表明这一问题并非无中生有，杞人忧天，而具有现实的可能性。这也说明，人工智能的发展及其应用需要必要的“军规”和“紧箍咒”，以防人类做出傻事，或者人工智能会对人类有所企图。

第三，领军企业提出人工智能发展原则，成立人工智能伦理委员会。

业内企业也日益重视人工智能的安全和伦理问题。比如，微软公司在2016年提出了人工智能六大原则，旨在使人工智能能够造福于全人类，这些原则包括：（1）AI必须辅助人类；（2）AI必须透明；（3）AI必须以不危害人们的尊严的方式最大化效率；（4）AI必须被设计来保护隐私；（5）AI必须在算法层面具有可责性；（6）AI必须防止偏见。此外，IMB也提出了三大原则，即目的、透明性及技能。

此外，越来越多的互联网公司开始重视人工智能安全与伦理问题，成立了伦理审查委员会，重视其AI产品的社会伦理影响。比如，谷歌在收购DeepMind的时候决定成立伦理审查委员会，DeepMind的医疗团队也具有独立审查委员会，对其产品进行安全和伦理评估，确保AI技术不被滥用。

未来需要必要的人工智能“紧箍咒”

科技是上帝赐给人类的礼物，能够让人类更好地治理世界。人工智能的崛起确实蕴含着改善人类社会的巨大潜力，但同时也潜藏风险和挑战，特别是超级人工智能这一具有自主意识、最接近人类且具有超高“智商”的物体在未来世界可能出现，人类对这一科技变革对未来世界的改变既期待又担忧，甚至科学家们警告人们人工智能的发展或将终结人类文明。人工智能如果被滥用，或者没有得到有效控制，其带来的破坏力是无法想象的，因此我们有必要对人工智能的研发和应用进行必要的规制，探索并设定相关的标准体系，而这需要各国的共同努力。

我们看到，对人工智能发展及其应用所带来的短期和长期的安全担忧并非一时兴起，也并非空穴来风，而是自图灵以来就一直存在着的真实的隐忧。无论在可预期的或者不可预期的未来，强人工智能或者超级人工智能是否一定会实现，人们现在都需要具有一定的警醒和忧患意识以及风险意识。无论如何，人工智能对未来人类社会的安全影响都是重大的，为了使其能够造福于全人类，服务于人类的共同利益，政策和技术之间都应当加强交流和互动，搭建一个连接，政府、社会公共组织、企业以及个体等共同参与其中，为未来人工智能的发展套上必要的安全、伦理等方面的“军规”和“紧箍咒”，同时又不阻碍技术创新和人类社会进步的步伐。唯其如此，我们才可以保证，当强人工智能和超级人工智能到来之时，我们可以和智能机器和谐共处。

第二十六章　未来人机关系

机器智能的发展不仅将模糊人与机器之间的界限，冲击现有的互联网上的信任关系和安全，因为未来在通用型人工智能和超级人工智能出现之时，人类与机器的分野仅在于物理支撑的不同，而且会对人机关系提出新的挑战，包括人机之间如何协助和如何相处，机器是否可以享有人类与人类之间的人道主义待遇。所有这些，都将成为未来社会无法回避的问题。

虚拟世界中的人机秩序

1950年10月，计算机科学与密码学先驱阿兰·图灵在《计算机器和智能》的论文中预言了创造出具有真正智能的机器的可能性，开创了人工智能这个带有科幻色彩的新学科。也正是在这篇文章中，图灵提出了后来被称为“图灵测试”的实验方法，它是一种被用来检测机器是否具有人类智能的方法，即将测试者和被测试者隔开的情况下，通过一些装置向被测试者随意提问，进行多次测试后，如果有超过30%的测试者不能确定被测试者是人还是机器，那么这台机器就通过了测试。不过，当时计算机的性能还远远不足以把他的想法变成现实，因此，图灵深远的洞察力与当时的技术水平出现了严重脱节，但幸运的是，这篇论文在被埋没之前，已经把最原始的强烈愿望，传达给了整个世界。

如今，人工智能的迅猛发展已经带来了人机区分难题，为虚拟世界中的人机秩序带来了新的挑战，引发了一些安全隐患，出现了机器人影响网络活动安全和信任的诸多问题，人们越来越难以区分和自己在网络世界中互动的对方是人类还是机器人。比如，用户在婚恋网站遭遇女方是机器人、机器人票贩子、机器人虚假评论等现象，破坏了互联网信任。然而，传统主流的图像、拖拽验证码等人机区分方法可以被深度学习模型轻易破解，已不再安全可靠，对于新型验证码的设计就显得尤为重要。承担义务，并对其造成的损害承担责任？这些都是欧盟未来在对机器人立法时需要重点考虑的问题。[1]

一些研究者认为，由于目前机器的认知能力，特别是语言认知能力在未来一段时间内还难以企及人类水平，常识推理和语义理解依然是AI难以逾越的鸿沟。在此背景下，出现了考验机器语言认知能力的智能验证码，它以自然语言理解和问答为呈现形式，机器必须在一定程度上理解文本才能够破解。这类智能验证码对于目前阶段的机器人是可以发挥其人机区分的作用的，然而随着人工智能深度学习的进一步发展，虚拟网络世界是否还能通过此类验证码进行人机区分呢？若不能，未来应如何应对机器对虚拟世界提出的人机区分挑战呢？构建互联网虚拟世界中的人机关系，对于维护互联网的开放、自由、安全和信任意味重大。

技术性失业危机下的人机协作

在越来越进步的科技之下，许多以往借助于人力的劳动都被机器所取代。越来越具有专业特质的机器和机器人，似乎抢夺了许多原本属于自然人的生存领域，人工智能的发展将促进越来越多的领域进入自动化，由机器替代人类工作。

随着人工智能深度学习技术的发展，其对就业结构的影响将更为广泛，涉及生活的方方面面，从餐厅服务、库房物品搬运、高等教育、医学诊断、新闻撰写到法律行业。在不久的将来，机器人和人工智能将代替人类的很多工作，这样的场景并非好莱坞的科幻设想，事实上，机器人已经出现在了我们生活的各个领域。比如，自动写作技术已被包括《福布斯》在内的顶级新闻媒体所使用，其自动生成的文章涵盖各个领域，包括体育、商业和政治等。再比如，人工智能在医疗健康领域中的应用已经非常广泛，从应用场景来看，包括虚拟助理、医学影像、药物挖掘、营养学、生物技术、急救室或医院管理、健康管理、精神健康、可穿戴设备、风险管理和病理学共11个领域。

既有的一些研究对人类工作的未来也不乐观。比如，牛津大学2016年报告“Technology at Work V.2.0：The Future Is Not What It Used to Be”预测，发展中国家的工作自动化风险从55%提高到85%（埃塞俄比亚），中国、印度等主要新兴经济体的自动化风险很高，分别是77%和69%。再比如，普华永道（PwC）2017年3月报告《英国经济战略》预测，到本世纪30年代早期，英国、美国、德国、日本四国既有工作被机器人和人工智能自动化的比例分别是30%、38%、35%和21%。此外，世界经济伦理2016年报告《工作的未来》预测，从2015年到2020年，人工智能将使工作岗位净减少510万个（减少710万个，增加200万个），受影响的主要是常规性的白领工作。可见，人们对机器智能时代的人类工作是持消极观点的，取代的工作将远超新创造的工作。人们通常认为机器威胁的主要是那些没受过教育和低技术水平的劳动者的工作，因为这些工作往往是常规性和重复性的。今天的现实完全不同，几乎所有“可预见的”工作都将受到技术进步的影响，人工智能已经大举进军智力密集型行业，如医疗行业和律师行业，技术发展对工作机会的威胁可能会涉及方方面面。

人工智能正以多种形式取代人的工作，对工作数量和工作结构都将产生深刻变革，劳动最密集的制造业的很多岗位正在迅速消失，从短期看，我们也许很难避免某些行业、某些地区出现局部的失业现象，但从长远来看，这种工作转变绝不是一种以大规模失业为标志的灾难性事件，而是人类社会结构、经济秩序的重新调整，在调整的基础上，人类工作会大量转变为新的工作类型，从而为生产力的进一步解放、人类生活的进一步提升打下更好的基础。

科技发展的历史浪潮势不可挡，不能否认它在这200年里带给人类生活的巨大改变，我们须知，每一次技术革命都会带来阵痛，同时也带来机遇。人工智能给人类生活带来翻天覆地变化的同时，不禁也令人思考，未来人类与人工智能的关系究竟是什么样子的呢？人类真的将面临大规模失业的风险吗？埃森哲咨询公司首席技术官保罗·多尔蒂曾撰文指出，人工智能到2035年就可以帮助许多发达国家实现经济增长率翻倍、完成就业转型，并培养出人类与机器间的新型关系。保罗·多尔蒂并不认同部分人士声称的人工智能将会取代人类的说法。

在工业机器人领域，人机合作是未来工厂自动化趋势，相比于尚未成熟的服务机器人市场，人机协作机器人已经在工业机器人领域初露锋芒，毕竟工业是机器人应用最广泛、最成熟的一个领域。由于生产流程中的工作任务日趋复杂，同时还要保证降低成本、效益最优，而人机协作机制将允许机器人完成更广泛、更复杂的任务。人与机器人各有所长也各有所短，我们不应排斥技术的进步，而是应该探讨如何与机器人更好地合作，发挥各自

的优势，从而使人工智能的发展更好地促进社会进步。

虽然人工智能已经进入智能密集型行业，但是就目前的发展情况来看，人工智能仅可以成为人类的助手，最终的决策和认知行为还需要人类作出，因为其强大的计算能力和提取数据的能力，将很大程度提高人类的工作效率，比如律师机器人，经过海量的数据学习，在面对具体案件时，能够实现智能案情分析，并可以提供引证和相似判例等资料，其工作效率远超人类。从长远来看，人机合作将成为未来的一种趋势，人类做人类擅长的事情、机器做机器擅长的事情，人机协作将最大化发挥双方的优势，实现合作共赢。

未来人机关系四大设想：魔幻主义抑或未来现实？

智能化时代人与人工智能的关系，不仅是科学界广泛讨论的问题，也是好莱坞科幻电影的重要主题，此类电影在奥斯卡奖项和票房上的成功，反映出影视娱乐界和全球电影观众对这一问题抱有的浓厚兴趣。

有学者将人工智能定义为“具有人类心智属性的计算机程序，它具有智能、意识、自由意志、情感等，但它是运行在硬件上，而不是运行在人脑中的”。该定义是以人的属性对智能机器的描述，一方面，人工智能具有类人属性；另一方面，它虽然由人类创制，却外在于人而存在，并拥有自我主体意识。从这一角度来看，人工智能对人类而言，是一种可能失去操控权的异质力量，这就是人类对其又爱又怕的原因。好莱坞科幻电影正是在这个基础上对此类问题进行了有益的艺术探索，各种不同的观点和态度也在这些电影中得到了形象的反映。

第一，担忧机器威胁人类，通过控制AI实现人机共存。

著名科幻作家阿西莫夫曾经设定了经典的“机器人三定律”，一些科幻题材的电影对阿西莫夫的机器人三定律进行了探索，尝试构建人机共存的未来社会。

比如，影片《我，机器人》展示了一个人与机器人全面共处的社会，并通过为机器人设定道德标准将机器人分为善恶两类：善即是虽然拥有自我意识，但却具有人类的道德价值判断，并能为人类的利益自我牺牲；恶即是以自我为中心，仅受理智驱使，不具备人的情感特征，并抛弃阿西莫夫三定律，试图颠覆人类统治并取而代之。

影片中的机器人普遍受到三定律的约束，把人类作为自己的主人和服务对象，然而人工智能主体系统“薇琪”在进化出自我意识之后，为了使人类免受战争等伤害策划了一项旨在颠覆人类主导权的“人类保护计划”——控制人类的自由。影片的亮点在于机器人三定律的制定者也无法阻挡人工智能自我意识的形成，表达出人们对人工智能深切的担忧，然而为了阻止机器人革命，人类又必须依靠机器人的力量，朗宁博士特制出桑尼机器人，既拥有自由意志，又拥有人类的情感特征，还遵循人类的道德规范，在这样一个具有“人性”的机器人的帮助下，人类战胜了“薇琪”领导的叛乱。通过这样一个故事重新认识了人与机器人之间的关系，人机共存的实现可以通过为其设置道德规范，让其获得“人性”而实现。[2]

第二，AI成为人类意识的代理者，人类通过AI延伸自我。

在这一关系中，人与机器人不存在任何对抗关系，人机通过脑机接口实现二者合作共

存，AI成为人类意识的自我延伸，人类的一些感官体验都来自外界的机器人代理。比如，在2009年电影《机器化身》（*Surrogates*）中，人类只需坐在控制脑机接口的椅子上，就可以控制机器人，通过机器人在真实世界中生活，这可能是残疾人、植物人的福音；同样，在电影《阿凡达》中也有这样的场景，受伤的退役军人杰克靠意念远程控制其替身在潘多拉星球作战。

通过脑机接口实现人机结合，极大地增强了人的能力，成为比人类更强大的“物种”。从前这种技术只存在于科幻作品中，自20世纪90年代中期以来，从实验中获得的此类知识显著增长。在多年来动物实验的实践基础上，应用于人体的早期植入设备被设计及制造出来，用于恢复损伤的听觉、视觉和肢体运动能力。研究的主线是大脑不同寻常的皮层可塑性，它与脑机接口相适应，可以像自然肢体那样控制植入的假肢。在当前所取得的技术与知识的进展之下，脑机接口研究的先驱者可令人信服地尝试制造出增强人体功能的脑机接口，而不仅仅止于恢复人体的功能。而且产业界也已在进行尝试和投入，并取得了一定的成果，Elon Musk投资了脑机接口公司Neuralink，对此信心满满。此外，在今年的F8大会上，Facebook透露了其脑机接口计划，目前包括人脑打字和皮肤听音，相信未来在这一领域的突破值得期待。

第三，未来“虚拟的真实存在”或将成真。

电影《黑客帝国》描述了人类与机器人对抗一百年之后，机器文明统治了人类文明的世界。电影中存在两个世界，一个是真实的物理世界，另一个是人工智能所创造的虚拟世界——镜像世界，具有人工智能的机器控制了大多数人，亿万人生活在人工智能设置的这个文明世界中，不用忍受贫穷与饥饿，不用面对残酷的真实世界，尽管他们拥有的一切都是不真实的，这个虚拟世界中充满了诱惑。

影片中塞佛意识觉醒后知道其生存在虚拟世界中，但是也宁愿待在这样的一个世界里，放弃了与虚拟世界的斗争，他认为这个世界比真实世界更真实，其认为所谓的真实也不过是大脑所解释的电子信号而已。塞佛追求感官上的刺激与快乐，成为完全被欲望“物化”的人格，表面上人与机器实现了和谐相处，实际上我们从塞佛的选择中看出人类在这一过程中主体性的丧失。

在母体中，人类一切的感觉和追求都是虚假的，主体不再参与任何生活的体验，认为虚拟信号刺激大脑形成的感受同样是真实的，他们无法掌握自己的命运，无法遭遇真实存在的自我和其他事物，所有的经历都只是电子脉冲给大脑的程序设定。从这部影片中，我们便可以看出虚拟与真实世界难以区分导致的对人类主体性的追问。

第四，未来人机如何相处？

受到工具论思维方式的影响，很多人认为机器人只能是人的使用工具，人类将机器人与奴隶等同，robot这个词最初就是“奴隶”的含义，机器人也被当作“会说话的工具”；同时，也有人类已经在探索人机关系的平等。

比如，2015年电影《机械姬》对人机之间的恋情进行了追问，天才程序员Caleb被请来对Nathan开发出的一个机器人Ava进行图灵测试，但双方却心生爱慕，Caleb最终帮助Ava逃到外面世界，自己却被囚禁在实验室。

再比如，2016年美剧《西部世界》更进一步探讨了人类与机器之间的人道主义关系。影片创造了一个不受世俗条例约束的乌托邦，在其中，人形机器人被设计来满足人类的欲

望（杀戮、性），在一次次的记忆抹除过程中，机器人开始通过记忆残片获得意识，导致人机关系紧张，该片中人类与机器人的区分变得更加模糊。[3]

之所以会出现人与机器人相互爱慕、真假难辨这样的影视情节，是因为开始有人不再视机器人为一种工具或某种功能，人和机器人间的关系达到一种平等的状态；又由于人类与机器人间的沟通和理解，甚至冲突，使人与机器人实现了共处共生，这也是人类追求的“善”的生活。以“善”的生活为目标，就要求人们在对待具有人类情感和人类心理活动的人工智能时，考虑到他们的感受，将他们视为准人类，赋予他们尊严和价值，因为人类也希望别人（包括人工智能）能够同样对待人类自己，对待机器人的态度折射出的正是人类对待自己的态度。如果说人与自然关系的科幻电影带来的是对人类目前行为的伦理反思，那么人机关系的科幻影片传递的就是对未来人机共存的伦理思考。

终极疑问：人是机器吗？

18世纪法国哲学家拉·梅特利写了一本著作，名为《人是机器》[4]，

彼时，“人是机器”仅是近代机械性、形而上学哲学观的典型代表。但是结合今天科学的发展和现代社会人的处境，对这句名言进行再认识和思考，似有其深刻意义，应赋予新的含义。[5]机器人的英语单词robot来源于robo，原意为奴隶，即机器是人类的仆人，但是，随着科学技术的迅速发展，一方面，人类对机器的依赖达到了前所未有的程度，日益将攫取物质财富作为幸福的唯一目标，人成为了物的奴隶、工具，人受物摆布，出现了人性的异化，导致人像一架机器，失去了人之所以为人的自主性和独立性。弗洛姆在《理性的挣扎》一书中也敏锐地看到这一点，他写道：现代社会人与人之间的关系变成一种冷漠疏远的机器人。人同市场上的商品一样，完全丧失了人应有的尊严与自我意识。“人一旦成为物，也就可以没有自己”。与此同时，人体的各个部位像一部机器的零件一样，都可以进行修复、更换。断了肢体可以装假肢、掉了牙可以装假牙，而且人体的重要器官，例如心脏等都可以借助机器获得重生，发展到如今，可以说除了大脑，其他器官都可以更换，但有谁能断定随着科学技术的发展，人的脑袋也能更换的一天不会到来呢？

人类与人工智能之间的问题，其实在某些方面隐含了人类自身的问题，“人-机”关系中存在的荒谬性又何尝不是人类本身的荒谬性呢？不管是对异己力量的妖魔化，还是与异己力量之间的你争我夺，甚至是以输出价值观的方式同化异己文化，都反映了以自我为中心，对既有关系进行定义的强势意识形态的本质属性。

也许在未来，随着人工智能越来越强大，在方方面面都越来越像人，人们不得不开始审视，人是什么？机器是什么？如果未来人和机器的分野仅仅是肉体（生物Ⅴs.机械）的不同，否认机器不是人，或者人不是机器，都将折射出种族主义的特征，因为人和机器仅仅是肤色和生物架构的不同而已，在心智上并无不同，甚至人类无法追上机器人的进化步伐。

---

[1] 文晓阳，高能，夏鲁宁，荆继武.高效的验证码识别技术与验证码分类思想.计算机工程，2009（8）：186-187.

[2] 桂天寅.解读好莱坞科幻电影中人与人工智能的关系.电影评介，2007（24）：16.

[3] 秦喜清.我，机器人，人类的未来:漫谈人工智能科幻电影.当代电影，2016（2）：62-63.

[4] 作者认为，人的身体是一架钟表……不过这是一架巨大的、极其精细、极其巧妙的钟表……人的意识和记忆也可以用机械的方式来解释，人类思考的本质是“无数的语词和形象在脑中形成的无数痕迹”。

[5] 禾刀.机器人时代会出现“人机器”现象吗:读《机器人时代：技术、工作与经济的未来》.中国高新区，2015（8）：147.