强人工智能体作为刑事责任主体的相关思考[1]

管亚盟

（中国政法大学刑事司法学院，北京 100088）

摘　要：基于人工智能技术的特征与人工智能体的特殊性，以过错为核心的一般侵权原则难以适用。产品责任也因“缺陷”要件难以认定和严格责任属性不断紧缩的趋势而出现规制困境，有必要讨论其独立承担责任的问题。人工智能体所造成的侵害基本可以达到评价刑事责任的程度，具有作为刑事责任主体的可能性。物理主义哲学对人工智能体自由意志的证成具有启发性，同时为了分散社会风险，将人工智能体拟制为刑事责任主体也具有一定的必要性。为实现刑事归责，应当尽快建立信息追索机制，构建人工智能体个体财产体系，同时尽快确立人工智能体伦理规则。

关键词：人工智能；自由意志；责任主体；风险分配

中图分类号：DF611　　**文献标识码：**A　　**文章编号：**

引言

随着人工智能技术的不断发展，我们逐渐进入强人工智能时代，即人工智能产品不仅具有工具性质，还可以独立思考，给人们思想上的启发。毫无疑问，强人工智能体的出现会改变人类的日常生活，但也带来了新的法学课题——“当强人工智能体给人类带来伤害，责任应当如何承担”。

一、讨论的必然性：产品责任的规制困境

对此，法学界有两大意见：第一，适用产品责任；第二，由人工智能体独立承担法律责任。笔者认为，只有在现有法律规范不能妥善处理强人工智能体的责任问题时，才要讨论新的方式，但目前来看，适用产品责任存在以下问题。

（一）人工智能体造成侵害的原因不能简单地视为“产品缺陷”

纵观世界各国关于产品责任的规定，该责任承担基本都是以“产品存在缺陷”为前提的，《侵权责任法》第 41 条关于产品责任的规定亦是如此。这便存在如何理解“产品缺陷”的问题。当前关于判断产品缺陷的通说理论是“产品投入流通时是否符合一般人合理期待的安全标准”，在具体操作过程中，则主要以是否符合国家和行业标准作为判断标尺。然而，目前尚未有国家和国际组织对人工智能体的生产设定相关标准，即便是在已经投入运营的自动驾驶汽车领域，号称具有里程碑意义的美国内华达州“511 法案”也未对相关技术标准作出直接规定，有学者指出，“由于无人驾驶汽车尚处于发展的初级阶段，受资金、人员配备等因素的影响，各国的研究进程和水平不一。即使是在同一国家，国内机构以及研究者的进程与水平不尽相同，这使得诸多技术尚无法制定统一标准”。因此，如何界定人工智能体的产

[1] 收稿日期：2019 年 12 月 9 日
作者简介：管亚盟（1993-），男，山东滨州人，在读博士，研究方向：刑法学。

品缺陷缺乏一个可靠的现实性标准，在具体操作上是极为不易的。

不过，最近有民法学者认为，“缺陷不等同于‘不符合质量标准’。对某些产品，并无国家标准或行业标准存在，但他们可能存在缺陷而无质量不合格问题；对于另一些产品，虽然不符合国家或行业质量标准，但并不存在缺陷”。由此，国家标准或行业标准在判断产品缺陷时并无实质意义，看似产品责任在规制人工智能侵害时又有了一丝曙光，实则却因人工智能体决策的独立性而否定了其作为“产品缺陷”的解释可能。

通常认为，人工智能是一门科学，目标是探索和理解人类智慧的奥秘[1]，而人工智能体则是通过机器将这种理解实现的有形载体。人工智能技术的发展从一开始便受到神经学的影响，科学家通过模拟人类思考的物理特点，利用计算机算法创造了一个“黑匣”，同时基于大数据平台，人工智能体可以不断自我学习，同时独立做出决策。由此，在准确认识人工智能技术的概念后，可以很清晰地认识到，人工智能体并非仅仅基于研发者的指令进行相关活动。“虽然学习算法可能是公开和透明的，但它产生的模型可能不是，因为机器学习模型的内部决策逻辑并不总是可以被理解的，即使对于程序员也是如此。”[2]正如人的决策过程具有不可解释性一样，为研究人的决策过程而诞生的人工智能同样具有这样的特征。因此，相比于一般产品，人工智能体存在一定的特殊性。仍以自动驾驶为例，研究人员仅仅制定了决策的基本规则和基本概念，好比驾校教练将驾驶机动车的基本技能传授于学员，而学员正式上路后，则有许多因素会干扰到他的具体决策，包括自身相关的素质。若发生交通事故，我们不能认为学员是驾校培养出的“缺陷产品”，类比到自动驾驶汽车和其研发者而言同样如此。

（二）严格产品责任的适用在世界范围内已逐步紧缩

多数学者认为产品责任可以规制人工智能体造成侵害的原因在于其是严格责任或无过错责任的属性，虽然对于此认识理论上也未达成一致观点[3]，但不可否认的是，严格产品责任的适用在全世界范围内已经逐步紧缩。

实际上，从产品责任适用的抗辩规则中可以看出，产品责任要求生产者至少存在一般意义上的过失。以发展风险抗辩为例，其起源于过失责任中的工艺水平抗辩，从理论上讲，过失的认定以行为人的预见可能性为基础，“若依据当时最新的科技水平都无法认识到该风险的存在，则自然意味着被告尽应有的注意也无法规避该风险”，[4]由此便可否定生产者的责任。20 世纪 60 年代，严格产品责任被奉为圭臬，发展风险抗辩被美国的许多法院拒绝使用。然而，严格产品责任的适用导致企业诉讼爆炸，经营面临困难。为了转嫁风险，企业抓住了商业保险这一根“救命稻草”，将增加的成本便转移到了消费者身上，造成产品价格大涨。在这样的背景下，巨额赔偿也让保险公司不堪重负，纷纷下架相关保险业务，导致行业萧条。这样的现状使得美国司法界开始反思，最终在 1998 年公布的《美国侵权法重述：产品责任（第三版）》中逐步接受了发展风险抗辩的相关原理，将严格产品责任仅限于“制造缺陷”的情形。而在欧洲，欧共体出台的《关于对有缺陷产品的指令》中也保留了发展风险抗辩的

相关条款，虽然同时规定了各成员国可以依据自身的情况选择排除适用，但从整体来看，目前在严格责任中保留发展风险抗辩制度的做法居于主流地位。在加拿大，法院基本不接受严格产品责任，原因在于：根据过失侵权索赔和根据严格侵权责任索赔的结果并无明显差别，已报道的案例中几乎没有产品被认定为有缺陷而生产者被认为为无过失的情形。[5]

人工智能技术的发展是一个全球命题，在发展风险抗辩事由逐步被各国各地区产品责任立法所接纳，严格产品责任适用逐渐紧缩的背景下，用产品责任规制人工智能体侵害行为的想法似乎难以实现。即便在我国，每次由于产品缺陷而爆发的大规模侵权事件，最终也都是通过一般侵权责任解决，产品责任并无适用的空间。综上所述，基于人工智能的特殊性，由其独立承担法律责任（主要是刑事责任）不失为一个解决问题的方法。

二、讨论的可能性：人工智能自由意志的证成与法律拟制

在古典主义所确立的道义责任下，刑事归责的前提是行为人具有自由意志，直到今天还在被坚持。虽然人工智能体可以在研发者所创造的规则下进行独立思考和判断，但其是否具有自由意志仍然是一个需要探讨的问题。

（一）关于人工智能自由意志存在的哲学思考

西方学者也对此问题进行了一番探讨。1950 年，阿兰·图灵（Alan Turing）在他著名的论文《计算机器与智能》（Computing Machinery and Intelligence）中首次提出了“图灵测试”（Turing Testing）的概念。该测试的内容是，让一个程序与一个测试人员进行五分钟的对话，然后由测试人员猜测是与程序或是与人对话，如果在 30%的时间内该程序愚弄了测试人员（让测试人员误认为是人类所答），那么这个程序就通过了“图灵测试”，也即可以认为它有自由意志。图灵还预言，到 2000 年，一台有 10^9 个储存单元的电脑可以很好地进行编程并通过“图灵测试”。[6]直到 2014 年，聊天程序“尤金·古斯特曼”（Eugene Goostman）通过了“图灵测试”，被认为是首次通过图灵测试的计算机软件。[7]

不过，“图灵测试”的理念在提出时就收到了许多哲学家的批评，他们认为，即便有机器通过了图灵测试，也不能说机器有了“实际”思考的能力，而仅仅是在“模仿”思考。不过，这一点图灵早就预见到了，他引用了杰弗里·杰斐逊（Geoffrey Jefferson）教授在一次演讲中的发言，“除非机器能够根据所感受到的思想和情感而写十四行诗或谱写协奏曲，而不是由于符号的偶然掉落，我们才能将机器和大脑等同起来——也就是说，机器不仅能写出来，而且知道它已经写出来了”[8]，并总结反对者的观点：机器必须有自我意识，才能认为它可以真正的思考。但图灵不以为然，它提出了一个反问：为什么我们要坚持对机器的标准高于人类的标准？毕竟，在日常生活中，我们从来没有任何直接证据证明人类的内在精神状态。[9]以此为开端，关于人工智能体是否可以“真正”思考，是否具有自由意识的争论在西方哲学界蔓延开来。

承接图灵关于反对者的反问，有哲学家开始反思，如果说人类有真实的思想，而机器没有，那么必须解决一个问题：即为什么人类的思想不是神经生理过程的结果？此即为物理主

义哲学家（Physicalist philosophers）关于自由意志的核心思想。为说明此问题，他们构思出了一个“钵中之脑”（the brain in a vat）的实验。大致是说，邪恶的科学家将人的大脑取出体外放入一个营养钵中，神经末梢与计算机相连，直到所有的指令都通过计算机发出，人的外部行为仍然不会改变。“我们可以设想，不只是一个大脑在钵中，相反，所有人类（或许所有有感觉的生物）之脑都在钵中。当然，那个邪恶的科学家必须在营养钵之外——要不他愿意吗？或许没有邪恶的科学家，或许这宇宙恰好就是管理一只充满大脑和神经系统的营养钵的一台自动机（尽管这是荒谬的）。”[10]因此，物理主义哲学家有理由认为，人的意志与计算机的意志一样，都是神经元物理过程的产物，不论这个神经元是天然的亦或是人工制造的。

当然，物理主义哲学家对自由意志的认识也受到了一些批评，最具有代表性的来源于美国哲学家约翰·希尔勒（John Searle）提出的“中文房间”（the Chinese Room）实验：一个房间中只有懂英文的人，房间里有用英文写的规则手册，以及各种各样的纸，有些是空白的，有些是让人无法理解的文字（中文），类推到人工智能，则书写规则代表了程序，人代表了CPU，成堆的纸代表了储存程序。小房间有一个对外的小开口，小开口处出现了纸条，上面写着难以辨认的符号，人可以在规则手册中找到指示，在房间中的纸堆里找到相应的符号，并进行排列，分析等。最终，这些指令的分析结果会转录到另一张纸上传出，但是，无论是输入还是输出，房间内的人都难以理解其含义。因此，设计正确的程序不一定产生理解，也即没有自由意志。

在特定的历史时期，哲学家们并没有见到强人工智能体的真实存在，因此，所有的讨论都是在假设中进行，存在一定的局限性。实际上，物理主义的观点与图灵最初的发问一直没有得到有效解决，甚至让一些科学家开始对人类自身的认识产生怀疑。“目前还没有一种被普遍接受的推理方式能从这些发现中得出结论，即拥有这些神经元的实体拥有任何特定的主观体验。这种解释上的差异导致一些哲学家得出这样的结论：人类根本无法对自己的意识形成正确的理解”。图灵本人也承认，意识问题是个难以解决的哲学问题。随后产生的哲学思考也未能跳出这一范式，直到今天也没能达成共识。

（二）物理主义对人工智能体自由意志的证成

关于“中文房间”的实验，即使排除随着实验的进程房间里的人通过学习可以逐渐摸清规律，进而对获取的资料和输出的文字产生理解（毕竟人工智能具有深度学习的能力）的可能性，仅以“不能产生理解”为由否定自由意志的存在，也是不能成立的。实际上，实验中的一幕在中国的乡土社会中便真实存在着。费孝通先生在其著作《乡土中国》中反对文字下乡的合理性，其最为核心的观点即是文字并不能在乡土性的社会中充当交流的工具，甚至对文化的传承都是没有用处的。关于前者，是因为在“面对面社群”里总会形成“特殊的语言”，“在每个特殊的生活团体中，必有他们特殊的语言，有许多别种语言所无法翻译的字句”[11]。关于后者，则是乡土社会因其强大的稳定性，“特殊的语言”足以传递世代间的经验，也即足够传承这个社群的文化，“时间的悠久是从谱系上说的，从每个人可能得到的经验说，确

实同一方式的反复重演”。抛开对文字下乡合理性的论证，单看乡土社会亦或在相似的社会秩序下生活的个体，最初他们并不会对自己所得到的经验和处理事情的方式产生理解，他们也仅仅是照着前辈所形成的规则去做出行为，进而发现收获了最好的效果，我们是否也可以否认这些个体的自由意志呢？显然不会。在“中文房间”实验所设定的情形中，房间里的人按照规则做出了行为，是因为这是规则下“最正确”的选择，虽然他们并不能理解。类比到人工智能的行为，虽然它们也不能理解行为本身，却依据规则做出了“最正确”的选择，实际与人类行为没有区别。

有的学者会反对，人之所以会被认为具有自由意志，是因为可以打破现有的规则而做出行为。然而在笔者看来，这实际上又落入了另一个规则之下，即“趋利避害”的功利主义。例如所谓的“看破红尘”，是因为身处“红尘”之中对其身体或是精神带来了折磨，因而选择逃避。正如“钵中之脑”实验中所描绘的，人看似在独立做出选择，却不自觉地到了同一个运行法则之下。近来，德国心理学家 Tobias Esch 研究发现，当人们因决策正确而获益时，大脑会向负责决策的区域发送“奖赏”信号，促进人的认知能力进一步提升，相反，当受到伤害时，杏仁核会因为“学会害怕”而产生恐惧记忆，让人不再做出相同的决策[12]。由此可见，“趋利避害”也不仅仅是人类情感上的选择，反而是由生理上的反应进行控制，落入了物理主义的论证规则之下。到今天，越来越多的科学结论开始证实人类情感与选择同身体的生理结构存在直接联系，甚至身体的改变会改变人原本的性情。笔者认为，物理主义虽然一经提出就饱受质疑，但至今没有一个新的理论能够将其完全否定，其自身是具有合理性的，可以作为认定人工智能体有自由意志的哲学依据。

（三）人工智能体也可通过拟制成为责任主体

通过研究可以发现，无论是我国学者关于人工智能体是否具有自由意志的论述，还是西方哲学对这一问题的思考，都在不自觉地将机器与人进行类比，这似乎有一个不言而喻的前提：人是有自由意志的。不过，我们也可以在前述图灵的质疑中发现，人也并未对自己的自由意志进行论证。“问题是，独立意思是推断或者拟制的产物”[13]。由此可见，人作为刑事责任主体并非具有当然性，也是法律拟制的结果。

纵观各国刑事立法，均对未成年人和精神病人有责任能力的限制。但从科学上讲，未成年人也有独立思考的能力，能够依据自己的行为做出选择；而精神病人，也按照自己的方式生活在自己的“小世界”里。理性主义和人文主义均对人之所以可以成为法律主体有了一些理论架构，在人工智能体能否成为刑事责任主体的讨论过程中，也有学者对此标准进行了相关介绍[14]，不过，很快就有学者敏锐地发现，“区分法律上人和非人的标准的问题就是何时需要戴面具、而且戴在谁身上、何时需要拆面具的问题，也就是判定某实体在何时、应否、可否被认可为法律主体的问题”，“法律人格是独立于任何物种具体特质的抽象存在”[15]。关于这一点，笔者是极为认同的。在经历了长途跋涉的论证后，似乎最后起作用的重要因素还是立法价值导向的问题。当然，建立在人工智能体有类人化的、拟制的自由意志的前提下，

这一切似乎更加顺理成章。

在笔者看来，通过拟制的自由意志（有被证实为真的可能性），可以将人工智能视为刑事法律责任主体。就当下的社会条件看，其最大的意义在于解放制造者和使用者的手脚，将风险由个体的承担转变为由社会进行消化。就这一点而言，在将人工智能体视为物的责任分配体系下是有理论障碍的。而在不远的将来，当超级人工智能体出现时，则可以保障人在规则制定中的优势地位，继续践行人工智能生而服务于人的价值理念，并为新的社会秩序的形成奠定基础。

三、讨论的实践性：人工智能体刑事责任的实现路径

（一）建立信息追索机制

人工智能体承担刑事责任要以主观具有过错为前提，这涉及故意、过失的判断。在司法实践中，对行为人主观状态的判断是通过客观行为进行推定的，对于人工智能体主观状态的判断也可适用这种方法。不过，相较于人而言，人工智能体的决策过程却是有“踪迹”可寻的。因此，有学者建议，可以在人工智能体上安装“黑匣子”，记录侵害的发生过程，以期明确责任主体[16]。笔者对此是认同的，但是，如果在发生侵害后出现黑匣损坏或丢失的情况，是否还有办法还原“事故”的全貌？答案是可以的，基于大数据的发展，人工智能的决策过程都可以上传至云端系统，但由云端系统的管理者对相关信息进行披露的可能性也带来了新的问题。

早在探讨网络服务提供者的披露义务时，就有学者提出过类似的担忧，网络服务商不提供相关用户信息，可能使得案件难以侦破，而提供相关信息则可能面临着违约带来的巨额赔偿。[17]这不仅仅涉及自然人用户个人隐私的保护，在人工智能体伦理规则确定后，通过何种程序，何种方式调取相关信息都是需要考虑的问题。在美国联邦立法中，网络服务商的责任豁免大致分为三类：首先是基于用户的同意；其次是提供给政府或者司法机关；最后则突出了对儿童权益的保护，可以“为了失踪和被虐待的儿童”而得到豁免。[18]不能忽视的是，虽然为政府或者司法机关披露相关信息可以追求责任豁免，但该部分立法仍然通过程序性规定严格限制公权力介入商业通讯信息的能力，详细规定了政府要求披露用户信息的事先通知、备份保存以及其他程序。[19]在我国，关于网络服务提供者的责任来源及豁免主要涉及网络言论管理及著作权保护的领域，涉及的法规有《全国人民代表大会常务委员会关于维护互联网安全的决定》第 7 条、《最高人民法院关于审理涉及计算机网络著作权纠纷案件适用法律若干问题的解释》第 4 条、《信息网络传播权保护条例》第 23 条等。《刑法修正案（九）》第 16 条和第 28 条分别涉及网络服务提供者披露信息和网络安全管理义务，我们可以推出司法机关调取证据是网络服务提供者违约责任豁免的主要依据，但目前尚无明确完整的规定规制公权力涉足相关信息的界限和程序问题。在人工智能时代来临之际，信息追索机制的建立有举足轻重的意义，随着科技的发展，信息追索的难度将逐渐降低，但隐私权的保护需要引起立法、执法及司法部门足够的重视，这应当被视为信息追索机制建立的重要环节。

（二）构建人工智能体个体财产体系

当前阶段，确立人工智能体法律责任主体的地位最主要的目标是要解决其致损后的赔偿问题，因此，需要构建人工智能体个体财产体系。在笔者看来，为了更好地分配风险，人工智能体的个体财产体系应当由生产者购买的强制保险、使用者因人工智能体所获得的部分收益以及国家成立的专项基金组成。

首先，关于生产者抑或设计者购买强制保险的方案，是最能为大众所认同的。毫无疑问，保险业的诞生顺应了风险分配的趋势，也为域外立法所采纳。例如，欧盟议会所辖的法律事务委员会针对人工智能体致损的立法建议中包含建立强制责任保险制度，要求智能机器人的生产者负有强制投保义务。[20]当然，购买强制保险投入设置不宜过高，理赔金额也不宜提出太高的要求，以保证生产者抑或设计者的持续发展能力，同时，保证保险业不会因此承担太大的理赔责任而导致业务萎缩。

其次，关于使用者因人工智能体所获得的部分收益，虽然没有真实的立法例作为参考，不过有学者建议，“法律规定，使用每个机器人的时候，必须向银行账户中转入一定的存款，万一发生机器人损害赔偿的时候，可以从中支取”[21]。这样的建议具有一定的启发性，然笔者认为，这只有在人工智能体不被人所有的情况下才能实现，而所有即包含了使用权能，我们没有理由要求所有者为使用人工智能体而支付两次对价。因此，需要在人工智能体的运营模式发生根本变革后，使用者所获收益才能作为人工智能体个体财产系统的一部分。例如，人工智能体被生产出来后即投向社会，不被任何人所有，使用者每次使用需在自己的收益中拿出一部分放入人工智能体的财产系统中，用于未来的责任承担。

最后，关于国家成立专项基金，国家收入最直接的来源是税收，每一个生活在国家中的合法公民都具有这项义务，所以这项举措是风险由社会承担最直接的体现。实际上，每当出现重大的社会风险难以化解时，国家基于公共管理的需要都要主动作为，这不单体现在对行业发展的规划中，也应当为行业风险的化解保驾护航。欧盟法律事务委员会也建议在强制保险难以覆盖的损害范围内设立赔偿基金，让投资者、生产者、消费者等多方主体参与这一机制，值得我们借鉴。

（三）尽快确立人工智能体伦理规则

如前所述，当前阶段人工智能体承担刑事责任的主要意义在于让社会分担其发展带来的合理风险，不过，随着人工智能技术的不断深入，由其是强人工智能体和超人工智能体出现的时候，刑事处罚则应当与道义责任相联系。2017 年 2 月，欧盟议会通过决议，考虑成立一个专门负责机器人和人工智能的欧盟机构，确立人工智能伦理准则；长期来看，考虑赋予复杂的自主机器人法律地位（即“电子人”）的可能性。2017 年 10 月，在沙特首都利雅得举行的“未来投资计划（Future Investment Initiative）”大会上，索菲娅被授予沙特公民身份，成为世界上首个获得公民身份的机器人。[22]由此可见，人工智能体在未来不仅仅可以作为刑事责任主体，甚至可能成为社会主体，学界应当尽快建立人工智能体相关的伦理规则，

不仅仅要强调人工智能体的义务，更重要的是赋予其相关权利。在笔者看来，这至少有三方面的积极意义：

第一，确立相应的刑罚措施。刘宪权教授曾提出，“建议增设能够适用于智能机器人的删除数据、修改程序、永久销毁等刑罚处罚方式，并在条件成熟时增设适用于智能机器人的财产刑或者权利刑等刑罚处罚方式”[23]。这已经涉及到机器行为的改变以及犯罪能力的剥夺，当然也有学者提出反对，“我们是希望通过法律来控制机器人的设计者、制造商以及使用者的行为层面，并且解决在机器人行为所产生的法律责任分担问题。法律不可能对机器人本身的行为层面产生影响，如不可能因为法律对机器人确立了过失赔偿责任制，就能提升其在行为时的‘注意水平’”[24]。笔者看来，确立人工智能体刑事责任主体地位，也应当为其配置刑罚措施，在理论上是一脉相承的。而反对者的观点也可以通过另一种解释方法予以化解：机器行为的改变需要相关技术人员的参与，但这绝对不是法律对技术人员产生直接影响，也即不是由技术人员直接承担具体的责任，而是利用他们的知识消除风险。退一步讲，刑罚对犯罪人的“注意水平”的影响又有多大？一般预防可遇不可[]，有学者认为，对于一般预防而言是可遇而不可求的事情，是“搂草打兔子”，应当“打的着就打，打不着就算”，搂草（实现刑罚的报应公正目的、特殊预防目的）是主要目的，能同时兼顾“打兔子”（一般预防）当然更好。笔者是极为赞同的。而特殊预防也需要矫正者煞费苦心，实际上，所谓的技术人员应当扮演矫正者的角色。反过来讲，在没有法律责任的情况下，技术人员不能随意对特定的人工智能体进行改变，这是以人工智能体有伦理属性为前提的。

第二，明确追责程序。明确责任主体后，紧接着需要研究的问题就是追责程序。目前，关于人工智能体刑事责任的探讨仅仅在实体法层面进行，根本原因就是人工智能体的伦理规则还未建立，不能明确其权利义务，更不能明确其在诉讼过程中的权利和义务，所以也无法对其设计相应的追责程序。如前所述，对人工智能主观状态（决策过程中保留）的判断是有迹可循的，这不仅设计用户的隐私，也涉及人工智能体自身的隐私问题。由此我们可以设想，追诉主体可以在多大限度内调取相关信息作为证据，以及非法证据能否作为裁判依据等现实问题仍然存在，人类世界的规则是否可以完全适用于人工智能体的追责程序，是一个值得思考的问题。

第三，培养公众认同。通说认为，刑罚的目的是兼具报应和预防的，之所以讲究报应，也因刑罚适用可以在一定程度上满足被害人的“复仇感”，起到抚慰其心灵的作用。但在人工智能体的伦理规则尚未建立之前，其仍然会被大众视为工具，“人们乘坐一辆无人驾驶的出租车，关心的是自己能否安全、便捷地到达目的地，而不是想和机器就出租车费用先‘讨价还价’一番”。此时，对工具适用刑罚并不能满足被害人的心理需求，可能会产生新的社会矛盾。虽然公众观念的转变需要一个时代发展的不断深入，但转变的前提是扭转人工智能体纯工具性的地位，这便需要伦理规则的建立。

参考文献：

[1]钟义信.人工智能：“热闹”背后的“门道”[J].科技导报，2016（7）：14.

[2]司晓，曹建峰.论人工智能的民事责任：以自动驾驶汽车和智能机器人为切入点[J].法律科学，2017（5）：170-171.

[3]周新军，容缨.论我国产品责任归责原则[J].政法论坛，2002（3）：68.

[4]贺琛.我国产品责任法中发展风险抗辩制度的反思与重构[J].法律科学，2016（3）：136-137.

[5]Goodridge v. Pfizer Canada Inc, [2010]O.J.NO.655;2010 ONSC 1095,p.40.

[6]A.M.Turing.Computing Machinery and Intelligence[J]，Mind，1950，（49）：433-460.

[7]刘春.计算机首次通过图灵测试，标志人工智能新阶段[EB/OL].（2014-06-09）[2018-09-28]. http://tech.163.com/14/0609/08/9U9KKHVV000915BD.html.

[8]Jefferson, G. The mind of mechanical man:The Lister Oration delivered at Royal College of Surgeons in England[J]. British Medical Journal, 1949，（1）：1110.

[9]Stuart J.Russell＆Peter Norvig . Artificial Intelligence :A Modern Approach[M]. Pearson Education,Inc.2010,1026-1033.

[10][美]希拉里·普特南.理性、真理与历史[M].童世骏，李光程译.上海：译文出版社，199：11-13.

[11]费孝通.乡土中国[M].北京：人民出版社，2015：15-22.

[12]德国心理学家：思维可改变大脑生理结构[EB/OL].（2017-04-10）[2018-09-30].http://www.sohu.com/a/132976687_710857.

[13]许中缘.论智能机器人的工具性人格[J].法学评论，2018（5）：159.

[14]龙文懋.人工智能法律主体地位的法哲学思考[J].法律科学，2018（5）：24-29.

[15]徐文.反思与优化：人工智能时代法律人格赋予标准论[J].西南民族大学学报：人文社会科学版，2018（7）：105.

[16]蔡婷婷.人工智能环境下刑法的完善及适用——以智能机器人和无人驾驶汽车为切入点[J].犯罪研究，2018（2）：26.

[17]秦天宁，张铭训.网络服务提供者不作为犯罪要素解构——基于“技术措施”的考察[J].中国刑事法杂志，2009（9）：42-47.

[18]18 U.S. Code § 2702.

[19]18 U.S. Code § 2703-2706，18 U.S. Code § 2516-2519.

[20]曹建峰.十大建议！看欧盟如何预测AI立法新趋势[J].机器人产业，2017（2）：18.

[21]储陈城.人工智能时代刑法归责的走向——以过失的归责间隙为中心的讨论[J].东方法学，2018（3）：30.

[22]心科技.解析世界首位被授予沙特公民身份的机器人——索菲亚[EB/OL].（2017-10-30）

[2018-10-8]. http://robot.ofweek.com/2017-10/ART-8321203-8440-30173530.html.
[23]刘宪权. 人工智能时代我国刑罚体系重构的法理基础[J]. 法律科学，2018（4）：47.
[24]吴习彧. 论人工智能的法律主体资格[J]. 浙江社会科学，2018（6）：62-63.

Reflection on Artificial Intelligence as the Subject of Criminal Responsibility

GUAN Yameng
（School of Criminal Justice, China University of Political Science and Law, Beijing 100088, China）

Abstract：Due to the characteristics of artificial intelligence technology (AI) and particularity of AI agent, the general tort principle with fault as the core is difficult to apply. Product liability also falls into a dilemma of regulation because "defect" element is difficult to identify and strict liability attribute is constricting constantly. So there is a need to discuss the independent responsibility of AI. The damages caused by AI agent can mainly reach the degree of being evaluated as criminal responsibility, and it has the possibility of being the subject of criminal responsibility. The philosophy of physicalism enlightens the justification of the free will of AI, and in order to disperse social risks, it is necessary to consider AI as the subject of criminal responsibility. In order to realize criminal imputation, we should establish information recourse mechanism, construct individual property system and set up ethical rules for AI as soon as possible.

Key words：artificial intelligence; free will; subject of responsibility; risk allocation

（责任编辑：于明明）