doi:10.20270/j.cnki.1674-117X.2025.1009

人工智能犯罪的人格构成及刑法回应

赵东，温栋木

（河北经贸大学 法学院，河北 石家庄　050061）

摘　要: 在本体视角下，人工智能通过计算机编程与建模、人工神经网络等技术手段，具备支配行为的“物化”意识人格要素；通过大数据的自我意识分析，具备基于因果功能联系的目的人格要素，其人格要素与自然人具有一致性。在社会视角下，人工智能在特定的行为领域中基于自身意识所作出的行为可以与特定的人与事物产生有限的联系，因而具备有限的社会人格要素，具有特殊性。因此，人工智能具备自我意识，能支配行为实现特定目的，进而在特定的行为领域具备完全的刑事责任能力。在此前提下，基于人工智能具备的有限社会人格要素，可创设强制学习劳动、销毁程序、缩减行为领域等针对人工智能犯罪的刑罚措施，以满足人工智能犯罪主体的刑罚适应要求。

关键词：人工智能；人格要素；刑罚；自主意识

中图分类号：D924.3　　**文献标志码：**A　　**文章编号：**1674-117X(2025)01-0082-08

Personality Constitution of Artificial Intelligence Crimes and Criminal Law Response

ZHAO Dong, WEN Dongmu

（Law School, Hebei University of Economics and Business, Shijiazhuang 050061, China）

Abstract: From an ontological perspective, artificial intelligence possesses the “materialized” consciousness and personality elements that govern behavior through computer programming and modeling, artificial neural networks, and other technical means. It also has purpose-oriented personality elements based on causal functional connections through self-conscious analysis of big data, which suggests that its personality elements are consistent with those of natural persons. From a social perspective, artificial intelligence, based on its own consciousness, can make limited connections with specific people and things in specific behavioral domains, thus possessing limited social personality elements and special characteristics. Therefore, artificial intelligence, with self-consciousness, can control behavior to achieve specific goals, thus having complete criminal responsibility in specific behavioral fields. Based on the limited social personality elements possessed by artificial intelligence, punishment measures for AI crimes, such as mandatory learning labor, destruction procedures, and reduction of behavioral fields can be created to meet the penalty adaptation requirements of AI crime subjects.

Keywords: artificial intelligence; personality elements; penalty; self-consciousness

收稿日期：2024-10-02

基金项目：河北省社会科学基金资助项目“人工智能犯罪刑事司法问题研究”（HB20FX012）

作者简介：赵东，男，河南信阳人，河北经贸大学副教授，博士，硕士生导师，河北省地方法治建设研究中心研究员，研究方向为刑法学。

现代刑法理论中，刑法的评价对象是人的行为，但不是所有的人所作出的行为都被纳入刑法考察的范畴，只有行为是人作出的且是在自由意志支配之下所产生的行为才是刑法所评价的对象。从这个意义上来说，刑法所评价的主体应当是具有自由意志的自然人。现行刑法中所规定的主体包括自然人与法人，但刑法对于法人的评价，其行为从本质上来看依旧是由自然人所作出的，公司的行为可以理解为自然人行为的集合。由此，现行刑法所评价的行为应当以自然人的人格作为基础底色，只有具备完整人格的主体所作出的行为才能被纳入刑法考察的范畴。从这个角度上来讲，关于人工智能犯罪是否应当由刑法来规制这一问题，就归结到了人工智能是否具备类似自然人的人格这一问题上。在人工智能出现之前，对于犯罪行为认定的研究，更多侧重于其是否在自由意志支配下作出的。对于动物来说，其不具备自由意志，所以动物的行为不是刑法所评价的具有人格意义的行为。而人工智能的出现对这一问题产生了冲击。人工智能是通过技术手段制造出人工的神经网络，并通过编程、建模以及仿真技术使其拥有自我意识，其行为是“可能超越人类设计和编制的程序在独立的意识和意志支配下实施的行为”[1]。以替代人从事生产活动为目的的人工智能，是人以自身为模板的创造物，天然具有“近人”“类人”的特点。人工智能“与人无限接近”的发展态势，使得人工智能作为法律主体的必要性与可行性问题在法学界引起持续讨论[2]。对于人工智能而言，判断人工智能的行为是否属于犯罪行为，就不能仅判断其是否具有自由意志，而需要在自由意志层面下更深入地探讨其是否具有如同自然人的人格。从表面上看，人工智能具有自由意志；但从更深层次来说，自由意志来源于完整人格要素。而将人工智能的行为纳入刑法评价，人工智能就需要具备刑法评价意义的人格要素。由此，就提出了人工智能是否具有人格、人工智能的人格与自然人有何不同等问题，这些问题也决定着人工智能的行为是否应当被纳入刑法的评价范畴。从行为理论的发展历程来看，从因果行为论到社会行为论再到人格行为论，在这些行为理论的发展过程中，对于行为的认定标准始终脱离不了人格要素，在每一个阶段都存在着对于主体人格的描述，且对于人格的研究也是随着理论的发展不断深入的。将这些理论中涉及人格要素的部分进行整理并深入探讨，可以为人工智能犯罪研究提供新的思路。

一、人工智能犯罪的人格要素分析

人格结构要素研究是随着行为研究的发展而不断深入的。在刑法行为理论研究中，虽然很少有关于人格结构的直接论述，但在行为理论发展的脉络中可以提取出有关人格要素的阐释。从行为理论阐释人格要素的视角来看，可以将其分为针对行为人本体视角的研究和整体社会视角下对行为的研究，其从本体要素和社会要素两方面阐释了什么是刑法意义上的行为。

（一）本体视角下人工智能人格要素的一致性

1. 本体视角下的人格要素结构

在行为本体视角中，最早的理论始于有意行为说。有意行为说中，意识这一要素首次被认为是行为概念的一部分，但是，关于意识的探讨依旧流于表面，对意识的理解依旧停留在生理结构的层面上，并未对意识要素进行深入研究，也未体现出意识背后的人格结构。在有意行为说之后，20 世纪 30 年代德国学者威尔哲尔又提出了目的行为论，目的行为论指出行为是人对目的的实现，在其看来，行为人与结果之间的目的关联才是刑法评价的行为的核心[3]。目的行为论认为，意识与行为的关系并非机械的反映，而是意识在理解现实因果关系的基础上，推导出行为对外界产生的影响以及可能会造成的结果，并以这种结果为导向产生一个目的，从而支配人为实现这个目的作出行为，人作出的行为就是为了实现其目的的外在表现。就目的行为论而言，意识以目的为导向引导行为人的行动，这是目的行为论认为的行为过程，这只说明了人作出一个行为时，其自身的意识与行为的关系。就是否具备刑法评价意义这一点来看，目的行为论认为在主体通过目的意识驱动行为达到结果的过程中，只需要具备目的与结果之间的联系即可。

有意行为说虽未说明意识与人格的具体内容，但在行为概念中确立了行为是受意志支配的自然人的举止行为[4]。可以说，在刑法评价的人格结构当中，意识是作为基础的人格要素。而在有意

行为说之后，目的行为论将意识这一要素作出了具体的解释，提出了由意识产生目的、由目的指引行为的行为人格结构。目的这一要素可以认为是人格要素结构中意识与行为之间的中介。从有意行为说对意识在行为结构中地位的确立，到目的行为论将意识这一要素具体化，二者都是在行为本体视角下进行分析，反映了刑法评价中作为基础人格要素的意识要素与作为中介人格要素的目的要素。

2. 人工智能与自然人犯罪人格要素的一致性

首先，就作为基础人格要素的意识要素来说，人的意识可以分为理性意识与感性意识。理性意识即人运用思维进行推理、想象，从而对外界反应的活动；感性认识是指人直接感受到的触觉、听觉、视觉，通过对外界的反应为人提供直接信息[5]。而人工智能则通过计算机编程与建模、人工神经网络、深度学习等模拟人的理性意识。通过光学成像、电子雷达以及大数据搜集来模拟人类的感性意识，再将通过由大数据、雷达等方法收集到的数据交给人工神经网络与程序进行分析处理，从而做到能以独立的自主意识去指引行为。人工智能通过技术手段，可以模拟人类的思维模式，再由这种类人的思维模式指引人工智能作出行为，在这个过程中人工智能若作出了犯罪行为，则其意识指引行为的过程与人类作出犯罪行为是相同的。正如约翰·罗杰斯·希尔勒所言，只要计算机运行适当的程序，它也是有思维的；只不过这种思维有别于人类的“程序思维”，一旦实现这种思维，人工智能体就能脱离人类的控制，在自我支配下依照自己的决策来行动[6]。人工智能虽然不具备人类的肉体载体，但从功能以及意识的运行方式上来说，这种“物化”的意识与自然人所具备的基础人格意识要素是一致的。

其次，就行为与意识的中介目的人格要素来说，目的之所以是意识要素与社会人格要素之间的中介人格要素，是因为目的要素是通过对因果关系的认识来预测行为和可能会导致的结果，并以这种结果为导向来指引行为的。一个人的意识与行为的关系是从基础的意识开始的，到最后行为反映其背后的社会人格特征，从而受到刑法评价，而目的的作用是作为基础意识与背后的社会人格之间链接的桥梁。易言之，目的行为要素是基础意识要素向社会人格要素发展的不可或缺的中介，而具有目的人格要素的基本体现是行为主体通过对因果关系的认识预测行为可能导致的结果。人工智能已经通过技术手段拥有了自主意识，在拥有自主意识的基础上，对人工智能进行数据喂养、智能感知、模拟训练等手段，经过自我意识的分析，人工智能可以认识现实世界中的因果关系这一链条，也能够通过对于因果关系的认识来指导行为，从而达到设定的结果。从这个意义上来说，人工智能的自主意识虽然是人工智能网络的编程与算法所体现出来的，但是人工智能的自主意识能够满足目的要素的基本环节。由此可见，无论是基础意识人格要素还是中介目的人格要素，在本体视角下，人工智能犯罪与自然人犯罪所具备的人格要素是一致的，人工智能能够通过技术手段具备个体意义上作出犯罪行为所需要的人格要素。

（二）社会视角下人工智能人格要素的特殊性

1. 社会视角下的人格要素

若仅仅从行为本体的视角对刑法评价中的人格要素进行分析，其结构无法适应刑法评价的社会属性。行为理论经过一定的发展，其研究重点也逐渐从行为的自然、生理要素转向行为具有的社会意义。在德国被普遍认可的社会行为说成功超越了在自然物理意义上对行为的研究，更加强调在整体的人类社会范围内对行为的认定。社会行为论认为，在讨论具有刑法意义的行为时，不能脱离刑法的功能这一要素，不能仅讨论行为本身的意义，而要将行为放在被刑法所规制的社会中去。研究犯罪行为是为了明确刑法针对的对象，从而更好地适用刑法，而适用刑法的目的是实现刑法的社会意义，其本质上是为了更好地利用刑法管理社会。从这个层面来说，对犯罪行为就有了多个角度的理解。首先，犯罪行为一定包含具有社会所承认的人的态度；其次，犯罪行为是在意识支配之下所作出的具有社会意义的行为，而行为是否具有这种社会意义是根据社会成员的一般认识亦即常识来进行判断并予以确定的[7]，是以一般人能够容易达到的事实性认识、理解为标准的[8]。易言之，犯罪行为是社会性的人的行为，人是身处社会中的人，而人的行为也应当是与社会具有相互联系的行为。社会行为论将行为放在

社会意义的框架中无疑是进步的；但是，什么是“社会意义”以及什么样的行为具有“社会意义”，其论述过于模糊，更多的是“把行为理解为价值关系的概念”[9]。由于对人格的具体内容没有论述，导致在对行为认定时，虽然可以解释作为与不作为、故意与过失，但在实践中不具有可操作性，导致适用上存在困难。

人格行为论指出人所作出的行为不过是自然人人格的外化，是主体内在人格在外部现实中的表现。团藤重光认为，行为本质是行为人主体人格的现实化[10]，人格行为论为行为赋予了完整的人格化意义。在人格行为论看来，人格与行为是互为表里的关系，行为人的人格通过行为表现出来，而行为又以人格为内在依据。行为人的意识以目的为导向操纵行为，这一过程是从行为人本体的角度来阐述行为的发生过程，要将行为放在社会范围内，就要继续探讨行为背后的社会性，而社会性应当体现在个体的先天因素以及后天所处的社会成长环境这二者共同塑造的具体人格之上。也就是说，作为刑法调整对象的行为，应当是由行为人的意识所支配的、以目的为导向并体现出其社会性人格的行为。人格行为论详细论证了具有评价意义的行为所需要具备的人格要素，不但将行为放在社会中观察，也强调了人格要素社会性的具体内容。

2. 人工智能犯罪人格要素的特殊性

从社会意义角度看，人工智能是否具备完整的社会意义上的人格要素？对于自然人来说，人具有双重属性，即自然属性与社会属性，具有刑法评价意义的是人的社会属性。换言之，只有在行为能够体现出人的社会属性的情况下，这个行为才具有被刑法评价的意义。而人的社会属性，本质上是人与社会及他人之间各种联系所构造的完整的社会人格，这种联系覆盖一个人的方方面面，包含着一个人与他所能接触的其他所有社会主体以及社会环境之间的联系。马克思在《关于费尔巴哈的提纲》中指出“人的本质不是单个人所固有的抽象物，在其现实性上，它是一切社会关系的总和”[11]。从更深层次上来说，具有社会人格意义的行为能够被刑法所评价，其针对的对象已经超越了行为本身，而转向其行为所体现出来的社会人格，只有具备社会人格才有刑法评价的意义。反观人工智能，虽然其在思维功能与结构上具备了人类作出行为时需要的意识要素，甚至通过一定的技术手段，在一些技术领域可以超越人类，但无论人工智能所具备的“智能”水平有多高，其终究无法像人类一样产生如此复杂的社会关系，也就不可能在人类社会中拥有与人类相同的社会属性。作为人类制造的客观产物，人工智能的意识也只能停留在自然物理层面上，其实施的犯罪行为背后无法体现出社会属性，也就不具备完整的社会人格要素。但是，在某些领域中，需要借助人工智能的技术水平来完成一些为人类提供便利的行为，例如医疗领域的人工智能手术以及人工智能汽车驾驶技术等，在这些领域人工智能虽不具备与整个社会复杂的联系，但在其行为领域内与被服务的人类可以产生单对单或单对多的小范围的联系。如果说，完整的社会人格要素是社会中相关的无数事物与人的联系所塑造的，那么对于人工智能来说，虽然不具备如同自然人一样复杂的完整的社会人格，但其在特定的行为领域中，人工智能通过本身的意识所作出的行为可以与特定的人与事物产生有限的联系，因此可以认为，人工智能具备有限的社会人格要素。

综上所述，刑法所评价的行为应当是人的意识支配的以目的为导向并能体现出行为人社会性人格的行为，在这个定义中包含了多个要素（见图1）。

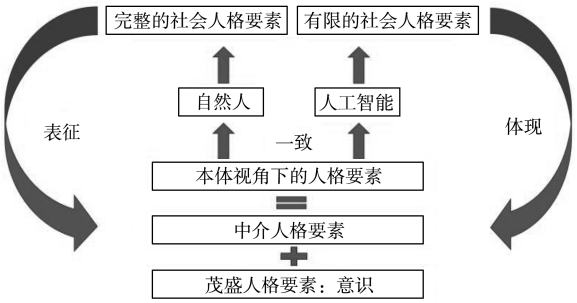

图1　刑法评价人格要素结构

如图1所示，想要满足刑法评价具有人格意义的行为，首先，从行为人本体视角来看，要具备作为行为基础人格要素的意识，以及通过意识对外界因果关系形成的认识，预测行为有可能导致的结果，并以此产生一个行为目的来指引行为，也就是具备目的人格要素；其次，在社会视角下，行为的目的要体现出背后隐藏的完整社会人格要素，这个完整社会人格要素的本质就是刑法评价

的具有人格意义的行为。对于人工智能犯罪来说，在本体视角下人工智能犯罪与自然人犯罪所具备的人格要素是一致的；而在社会视角下，人工智能虽然不具备完整的社会人格要素，但在特定的领域内，人工智能通过其行为可以在有限的空间内与社会产生一定的联系。也就是说，人工智能虽然不能像自然人一样拥有复杂的社会关系，但其也并不是完全处于社会主体的联系之外，在人工智能具备了社会人格要素的前置条件，即本体视角下的意识要素与目的要素之后，行为又与社会产生有限的联系。我们可以认为，人工智能的行为背后存在着有限的社会联系，即人工智能具备有限的社会人格要素。

二、人工智能犯罪人格的刑法评价

（一）人工智能犯罪人格的行为认定

行为是刑法评价最直观的要素，对行为的认定具有重要意义。就人格行为论而言，自然人所作出的犯罪行为必须体现出其背后的人格意义才能被认定为犯罪行为。行为并不是单纯的身体动静，只有行为主体在作出身体动静时，此身体动静与行为主体的人格相统一，达到行为是人格的外化、主体人格是行为的体现，在这种情况下，其才被认定为行为。例如，就过失行为来看，虽然过失行为并不是行为主体有意作出的身体动静，但人之所以会作出过失的行为，是基于其主体人格对规范的蔑视，所以过失是行为。不作为也能体现人格态度，不作为是指行为人负有作为的义务，其能够作为而不作为从而导致结果发生的行为；从客观行为上来看，行为人并没有作出任何积极的身体举动，但其不作出应做的行为，可以体现出其人格对规范的否定态度，所以不作为也是行为。反过来说，一些不能反映主体人格的行为，例如梦游或单纯的神经反射，仅仅是单纯的身体动静的行为不是刑法意义上的行为。如前所述，人工智能虽然不具备行为完整人格要素，但在特定领域中具备有限的行为人格要素，也就是说，在特定的情况下，人工智能的行为能够体现背后的有限的社会人格要素。

从具体行为来看，可以将人工智能相关的犯罪行为分为三种类型：第一，自然人利用人工智能进行犯罪的，这是指犯罪行为并不是出于人工智能本身的自主意识，而是自然人将人工智能当作工具并作出错误的引导或指令，导致人工智能作出危害行为；第二，人工智能被设计之初就是为了实施犯罪行为或因为设计缺陷导致实施犯罪行为；第三，人工智能在设计之初没有缺陷，也没有人恶意利用，是人工智能在自主学习的过程中产生了实施犯罪行为的主观态度，从而实施的犯罪行为。就以上三种类型而言，笔者认为第一种类型与第二种类型不能认定为人工智能所作出的犯罪行为，人工智能所作出的犯罪行为应当是基于人工智能的有限人格要素的外化。第一种类型应当被认定为自然人犯罪的间接正犯，即自然人利用人工智能的不知情或通过一定手段绕开人工智能的自由意志，引导或支配人工智能的行为从而实施犯罪，类似于利用不知情的自然人实施犯罪，其行为的作出并不是出于人工智能本身的目的要素，不能体现出人工智能的人格意义。在第二种类型中，人工智能在设计之初就被人为植入了已定的程序或者设计本身存在缺陷，这导致人工智能在诞生之初就存在行为偏向，在这种情况下，要么人工智能缺少完整的意识人格要素，要么存在已经被决定了的出于非自身意识的目的，无论是哪种情况，其本体人格要素都不完整，不能认为是完整意义上的拥有自由意志的人工智能，应当追究设计者的责任。而第三种类型，人工智能是完全基于自身的意识要素与目的要素所作出的行为，只有在这种情况下的行为才能被认为是人工智能作为本体作出的行为。

认定人工智能的行为，除了需要考虑上述行为类型外，还需体现其社会人格要素。人工智能只在特定领域中具有有限的社会人格要素，这里的特定领域指的是被社会所认可和信赖的、可以交给人工智能进行工作的领域，只有在这种领域下才能认为行为与社会产生的联系是体现人工智能的人格要素。例如，投入使用的具有生产合格许可的人工智能驾驶的汽车，其在公共交通领域是受到社会的许可与信赖的，此时因对社会的工作服务关系，人工智能被赋予了一定的权利与责任，基于这种权利与责任，人工智能的行为与社会产生了一定的联系，这时人工智能若因为自主意识实施犯罪行为，其行为体现的是人工智能本身的有限社会人格要素，因而人工智能的行为能够体

现人格要素，是具有刑法评价意义的行为。反过来说，不被社会认可与信赖的人工智能，无论是出于何种原因，人工智能在不被信赖、未投入使用的情况下都不能与社会产生一定的联系，这种不信赖的本质是对设计者的不信赖。在此情况下，应当严格限制人工智能的行为限度，并对其行为进行严格监管，若人工智能这时作出犯罪行为，不能认为是人工智能的犯罪行为，其所反映的是监管者的社会人格，应当追究监管者的故意或过失的责任。

（二）人工智能犯罪主体的刑事责任能力

刑事责任能力是指进行责任非难所要求的行为人的能力，行为主体不具备有责地实施行为的能力时，不能对该行为进行法的非难[12]。判断主体是否具有责任能力，在刑法理论中有新派与旧派之分。旧派认为，责任能力是行为人通过自由意志支配自己行为的能力，即有责行为能力、意思能力、犯罪能力，其本质上是自由意志的问题。行为主体在具有自由意志的情况下是有能力选择自己的行为是向“善”还是向“恶”的，在认识到“善”与“恶”时，却决意向“恶”，那么行为主体就应当承担道义上的责任。新派认为，刑事责任的本质是对刑罚的适应能力，即行为主体的社会人格是否能被刑罚改造。但是，按照新派的理解，那些无法被刑罚所改造的主体，如累犯，会被认定为无刑事责任能力的主体，这是不妥当的。可罚的责任说认为，旧派与新派关于刑事责任能力本质的理解并不是二元对立的，二者相互之间存在关联。要求行为主体具有自由意志，能够选择自身的行为走向，也是为了对行为主体适用刑罚。某些情况下，行为主体在作出犯罪行为时具有自由意志，从旧派的角度应当承担责任，但从政策角度来看对其科处刑罚是不必要的或是会造成重大的恶劣影响，即在行为主体不具备刑罚适应能力但具有自由意志的情况下，可罚责任说也不认为其具有刑事责任能力。如前所述，仅坚持旧派会导致那些不具有刑罚适应性的主体被科处刑罚，浪费社会资源，而新派则又无法解释累犯等现象。因此，可罚责任说应当得到支持。

人工智能是否具有刑事责任能力需要从两方面进行考察。首先，人工智能通过技术手段可以具备自然人的意识以及目的，虽其不具备完整的社会人格要素，但可以认为人工智能具有可选择性的自由意志。其次，人工智能是否具有刑罚适应能力？现有的刑罚种类，主要是剥夺犯罪人的人身自由或生命的自由刑和生命刑，这样的刑罚对于人工智能体而言能否产生“剥夺的痛苦”，从而起到刑罚的效果？答案显然是否定的[13]。这一类的刑罚是针对自然人所设立的，人工智能对此不具备适应性。但就刑罚的目的来看，若创设针对人工智能的刑罚种类，在人工智能主体上实现刑罚的目的是可行的。因此，人工智能可以同时满足有责行为能力与刑罚适应能力，人工智能具有刑事责任能力。

（三）人工智能犯罪人格的刑罚回应

1. 刑罚评价本质

人工智能犯罪与自然人犯罪从人格结构上来说存在一定区别，人工智能的犯罪行为虽然存在空间上的限制，但可以承认人工智能的行为能够构成犯罪。人工智能的行为人格结构也决定了其具有一定的刑事责任能力。刑罚是犯罪的直接法律后果，其内容是国家给犯罪人施加某种痛苦、折磨，使其遭受一定损失或丧失某种社会地位。报应刑论认为，之所以要对犯罪主体施加刑罚，这是对其作出的犯罪行为所产生的恶报，即犯罪主体作出“不正义”的行为就应当得到针对其个体的“不正义”的结果，就是作为恶报的刑罚[14]。犯罪行为本质是犯罪主体社会人格的体现，而刑法所评价的对象实际上也是犯罪主体的社会人格要素，从这个意义上来说，刑罚对犯罪行为所应施加的“恶报”，其本质上应当是针对犯罪主体人格的否定评价。而目的刑论认为，刑罚本身是没有意义的，对犯罪主体施加刑罚这一方法，只有在能够达到刑罚的目的时才能将刑罚正当化[15]。刑罚的目的包括一般预防与特殊预防，特殊预防是为了防止犯罪人再次犯罪所施加的刑罚，也就是说，犯罪主体作出一定的犯罪行为后，刑法认定其人格不适格，于是对其施加刑罚予以改造以防止其再次实施犯罪行为，其本质是刑法在其作出犯罪行为后，对行为背后所体现出的人格的否定性评价。由此可见，无论是报应刑论还是目的刑论，都可以得到相同的结论，即刑罚对行为施加的惩戒只是刑罚的外在形态，其实质内容是对支配行为人行为背后人格的一种否定性评价。

2. 人工智能犯罪人格的刑罚应对措施

我国现行的刑罚种类有自由刑、财产刑、资格刑和生命刑，这些刑罚种类是针对自然人而设置的。因为人工智能与人类的自然存在方式与社会人格内容的不同，针对自然人所建立的刑罚体系对人工智能并不适用。就上述四种刑罚种类来看，首先，人工智能是基于编程和人工神经网络而存在的，其物理载体是电子计算机，人工智能不依存于人的身体，也就不存在人身自由与生命的概念，所以自由刑与生命刑并不适用。其次，虽然人工智能具备有限的社会人格，但其与社会的联系仅限定于人工智能工作的领域，不能随意与社会建立联系，人工智能并不享有完整的社会主体资格，也就不能参与政治生活，剥夺政治权利的资格刑对人工智能也无法适用。最后，人工智能不具独立财产的权利，人工智能不拥有财产权，财产刑对人工智能也没有意义。人工智能不同于自然人和单位，其既没有生命、人身自由，也没有财产（就目前阶段而言），所以其既不能适用生命刑、自由刑，也不能适用财产刑 [16]。

刑罚的目的包括特殊预防与一般预防，由于人工智能不具有完整的社会人格，对单一的人工智能实施刑罚无法对其他人工智能主体产生警诫，所以人工智能的刑罚措施应当以特殊预防为主。根据人工智能的特点，可以构建适用于人工智能的刑罚措施体系，笔者认为，以强制学习劳动、销毁程序、缩减行为领域作为人工智能犯罪的刑罚措施较为妥当。人工智能相较于人类更加容易实现刑罚特殊预防的目的。行为是社会人格的外化表现，人工智能的人格载体是电子计算机，针对人类设计的单纯以时间计算的自由刑对于人工智能没有实际意义，但就刑罚的目的来看，现有的自由刑，尤其是有期徒刑的执行中，包含着对自然人罪犯强制劳动教育改造的内容。虽然自由刑的时间意义对于人工智能并无价值，但教育改造的目的可以通过学习劳动实现于人工智能，因此可对人工智能设立强制学习劳动的刑罚措施，通过大数据喂养学习、劳动实践等措施改造人工智能，达到刑罚特殊预防的目的。而对于已经无法通过劳动教育改造从而回归社会生产生活的人工智能，可以对其设立销毁程序的刑罚措施，从物理层面消除其存在，从而达到特殊预防的目的，即相当于对于自然人罪犯的生命刑。以上两种刑罚措施是针对人工智能本身的人格要素达到刑罚特殊预防的目的，应当作为人工智能的主刑存在，实施强制劳动学习的同时，还应根据人工智能的数据情况与对数据的改造情况，对人工智能实施缩减行为领域的措施。这里的活动领域缩减包括禁止人工智能在某个领域中工作以及在某个特定领域内减少人工智能的行为限度，在一定期限内进行考察，若强制劳动学习后的一定时间内人工智能无违规行为，可以适当拓宽其工作领域。

3. 人工智能犯罪人格的刑罚效果评价指标

针对自然人的刑罚方法对人工智能并不适用，但其刑罚的本质是相同的。针对自然人的刑罚种类，如资格刑、财产刑、自由刑、生命刑，都是人类作出犯罪行为后，为了预防犯罪对行为人施加的造成其一定损失的措施，这些刑罚的本质是对自然人人格的否定。对于人工智能实施刑罚的本质同样是刑法对人工智能人格的否定性评价。强制学习劳动可以理解为在人工智能作出犯罪行为后，刑法否定其行为背后体现的人格要素，通过强制学习劳动的方式排除其人格中的否定要素，从而回归社会生产生活。销毁程序则是对人工智能人格要素的整体否定。这里需要说明的是，销毁程序与死刑不同，在实施销毁程序刑罚后，人工智能的现实载体并不会被销毁，被销毁的是人工智能的程序与数据，即人工智能的人格要素，而人工智能的现实载体依旧可以再次应用。缩减行为领域这一刑罚措施是对人工智能有限的社会人格的否定。人工智能在其特定的工作领域内可以与社会产生有限的联系，这种有限的联系是人工智能主体社会人格的基础。缩减行为领域通过减少人工智能的工作以及行为的范围，减少其与社会的联系，从而达到对有限社会人格的否定。对人工智能犯罪的刑罚效果评价指标，应当集中于对人工智能社会人格的正确改造。人工智能的社会人格包含了意识、目的、与社会联系这几个要素，人工智能通过不正确的自主学习，塑造了不适格的社会人格，进而实施犯罪行为。通过对人工智能实施刑罚，将人工智能的意识、目的与社会联系进行修正，改造其社会人格，从而在社会生产生活中排除具有不适格社会人格的人工智能主体，保障社会安全，这是评价人工智能刑罚

效果的依据。

社会发展日新月异，人工智能的进化速度超出人类预期，人工智能逐渐参与到我们的社会生产生活中。在传统理论中，刑法意义上的行为主体，无论是自然人还是法人，其本质都是基于人的肉体所作出的行为；当人工智能这种不依存于人的肉体且拥有自主意识的主体出现时，传统理论难以界定人工智能能否成为犯罪主体，现有刑罚措施也并不适用于人工智能这类特殊主体。虽然人工智能已参与社会生活、拥有自主意识，但其与人类在社会中的主体地位仍存在显著差异，因此，我们需要通过对人工智能主体人格要素的深入研究，来探讨其作为犯罪主体的理论依据，并探索适用于人工智能主体的刑罚措施。

参考文献：

[1] 刘宪权 . 人工智能时代刑事责任与刑罚体系的重构 [J]. 政治与法律，2018(3)：89.

[2] 王瑞玲，周月 . 赋予人工智能民事主体资格肯定论 [J]. 湖南工业大学学报（社会科学版），2023，28(5)：61-68，77.

[3] 许玉秀 . 当代刑法思潮 [M]. 北京：中国民主法制出版社，2005：99.

[4] 克劳斯·罗克辛 . 德国刑法学总论：第 1 卷 [M]. 王世洲，译 . 北京：法律出版社，2005：247.

[5] 玛格丽特·博登 . 人工智能的本质与未来 [M]. 孙诗惠，译，北京：中国人民大学出版社，2017：56.

[6] SEARLE J R. Menti, Cervellie Programmi[J]. Un Dibattito Sull’ Intelligenza Artificiale，1984(1)：1-22.

[7] 野村稔 . 刑法总论 [M]. 全理其，何力，译 . 北京：法律出版社，2001：269.

[8] 大塚仁 . 犯罪论的基本问题 [M]. 冯军，译 . 北京：中国政法大学出版社，1993：89.

[9] 杜里奥·帕多瓦尼 . 意大利刑法学原理 [M]. 陈忠林，译 . 北京：法律出版社，1998：105-106.

[10] 大塚仁 . 刑法概说：总论 [M]. 冯军，译 . 3 版 . 北京：中国人民大学出版社，2003：113.

[11] 李爱华 . 马克思主义经典著作导读 [M]. 北京：北京师范大学出版社，2008：8.

[12] 张明楷 . 外国刑法纲要 [M]. 3 版 . 北京：法律出版社，2020：170.

[13] 吴宗宪，张进帅 . 生成式人工智能：风险、挑战与刑事规制：以 ChatGPT 为例 [J]. 中国特色社会主义研究，2024(3)：61-72，92.

[14] 马克昌 . 近代西方刑法学说史略 [M]. 北京：中国检察出版社，1996：508.

[15] 马克昌 . 比较刑法原理 [M]. 武昌：武汉大学出版社，2002：752.

[16] 刘宪权，朱彦 . 人工智能时代对传统刑法理论的挑战 [J] 上海政法学院学报（法治论丛），2018，33(2)：44-51.

责任编辑：徐海燕