中外法学 Peking University Law Journal
Vol. 36, No. 4(2024)pp. 884－898

人工智能伦理的法律性质

李学尧*

摘 要 将人工智能伦理从道德原则转化为可操作、可预期、可计算的伦理合规实践，需要探究人工智能伦理的法律性质，特别是法律体系如何评价以及纳入人工智能伦理规范。可以用“法化”和“事物本质”这两个传统法社会学理论曾用以反思规范理论的分析工具。一是，相比于“法治化”“规范化”等被泛化的概念，“法化”用以分析人工智能伦理的实操问题，具有更好的理论收敛性。科技伦理的既有法化路径主要有三：道义论持有者常用的新兴权利证立（如人格权）、结果论持有者常用的软法化（如政府监管工具创新）、美德论持有者常用的共同体伦理（如职业伦理）。二是，立法者或者“法律发现者”，在试图从人工智能技术及其产业的特征中推导出规范性的伦理要求的过程中，可以将事物本质的理念纳入，通过与生物医药伦理比较，抽象出人工智能伦理法化的三个约束性要件：道德规则的技术可嵌入性、更强的场景性以及依赖于技术过程的程序性。

关 键 词 人工智能伦理 科技伦理 伦理审查 法化 事物本质

伦理被认为是人工智能治理的重要手段，也是当下相关立法研讨的焦点话题之一。科技部等单位2023年颁布的《科技伦理审查办法（试行）》（以下简称《伦理审查办法》）第2条及其附录“需要开展伦理审查复核的科技活动清单”，在准立法意义上构筑了人工智能伦理审查的专门制度。〔1〕《伦理审查办法》颁布以后，人工智能产业界及其合规服务群体有较多意见，〔2〕主

* 上海交通大学凯原法学院教授。本文系国家社科基金重点项目“支持全面创新的基础法律制度研究”（项目编号：22AFX003）的阶段性研究成果。

〔1〕 2022年国家网信办颁布的《互联网信息服务算法推荐管理规定》是我国最早提出人工智能伦理审查要求的法律文本。该规定第7条规定“算法推荐服务提供者……建立健全算法机制机理审核、科技伦理审查、用户注册、信息发布审核、数据安全和个人信息保护、反电信网络诈骗、安全评估监测、安全事件应急处置等管理制度和技术措施”。但对于如何开展伦理审查，该规定并未展开。

〔2〕 类似思路的学术表达可见孟令宇、王迎春：“探索人工智能伦理审查新范式”，《科学与社会》2023年第13期，第97—113页。

要表现在:《伦理审查办法》是基于生物医药伦理的思路起草的,未充分考虑人工智能伦理的特征,因此在具体的合规过程中会给研发者和相关企业带来实操性困扰。我们可以综合运用来自法社会学传统的"法化"概念和来自法学方法论传统的"事物本质"两个概念,〔3〕通过分析人工智能伦理与生物医药伦理、职业伦理的区别,面向人工智能的立法实践需求,呈现人工智能伦理的法律性质,特别是法律体系如何评价以及纳入人工智能伦理规范,以期待为相关的人工智能伦理制度构建及其实操化,提供系统性的理论参考。

一、为什么采用法化的思考路径?

(一)从道德原则框架到伦理合规实践的转化需求

在应用伦理学内部,相比于生物伦理学、商业伦理学、职业伦理学等,人工智能伦理的相关研究似乎是个新事物,但它的理论可以上溯至上世纪六十年代,还略早于生物伦理学的产生。后者作为一门学科,一般被认为产生于20世纪七十年代。〔4〕1960年,《科学》杂志上展开了有关于自动化带来的伦理问题讨论。〔5〕此后,网络信息、区块链、大数据、元宇宙等各种前沿信息技术的大规模社会化应用也伴随而来相应的伦理讨论热潮,出现了"信息伦理""信息技术伦理""大数据伦理"等类似的理论概念及其制度实践。〔6〕

随着人工智能技术及其应用场景越来越复杂,特别是越来越有能力执行更复杂的人类任务,它们的行为变得更加难以监控、验证、预测和解释,因此各国政府、各类国际组织、各类专业组织和相关平台企业在近十年里就人工智能伦理的重要性及其规则内容提出各种各样的倡议、原则、准则和指南。〔7〕对于它的争论,已经从最常见的数据安全治理(包括数据财产权益保护、隐私保护、知情同意、准确性和深度造假),逐渐扩展到社会公平(包括分配正义、失业、性

〔3〕"事物本质"虽是法律方法论中的一种理论工具,似乎与自然法学、法教义学传统更接近,但实质上它也常被法社会学理论用来"反抗"规范理论。参见舒国滢:《战后德国法哲学的发展路向》,《比较法研究》1995年第4期,第337—355页。

〔4〕See A. R. Jonsen, "A History of Bioethics as Discipline and Discourse," in Nancy Ann Silbergeld Jecker, Albert R. Jonsen and Robert A. PearlmanNancy Ann Silbergeld Jecke (eds.), *Bioethics: An Introduction to the History, Methods, and Practice*, London: Jones & Bartlett Learning, 2012. pp. 3-16.

〔5〕See Norbert Wiener, "Some Moral and Technical Consequences of Automation," *Science*, Vol. 131, No. 3410, 1960, pp. 1355-1358; Arthur L. Samuel,"Some Moraland Technical Consequences of Automation: A Refutation," *Science*, Vol. 131, No. 3410, 1960, pp. 1355-1358.

〔6〕See Kord Davis, *Ethics of Big Data-Balancing Risk and Innovation*, Sebastopol: O'Reily Media, 2012; Adriano Fabris, *Ethics of Information and Communication Technologies*, Ebook: Springer, 2018. 中文的文献综述可以参见邱任宗、黄雯、翟晓梅:"大数据技术的伦理问题",《科学与社会》2014年第1期,第36—48页。

〔7〕法学视角的综述可参见宋华琳:"法治视野下的人工智能伦理规范建构",《数字法治》2023年第6期,第1—9页。

别歧视、种族主义等),甚至还包括人工智能体的法律主体地位等。〔8〕2022 年生成式人工智能进入公众视野之后,社会各界愈加重视人工智能伦理的治理功能。〔9〕

在全球范围里,人工智能伦理治理实务中普遍存在注重原则层次的价值宣言但缺乏执行力度的问题。〔10〕欧盟可能是最早发布具有"法化"意义的专门性人工智能伦理规则的法域,早在 2019 年 4 月 8 日就通过欧盟委员会发布了《人工智能伦理准则》,提出"可信赖人工智能"的定义及相关伦理要求,但在合规实践中也给业界带来了"难以操作"困扰的批判。〔11〕在 2019 年,中国也发布了《人工智能北京共识》,但这不同于欧盟《人工智能伦理准则》,后者是官方机构发布的,而前者主要是高校院所和产业联盟。即使在欧盟 2023 年 12 月发布的《人工智能法案》中,也只针对高风险人工智能系统原则性提出强制性遵守的伦理规则,它实质上与以往"去道德化"的法律规则无异,说它是人工智能伦理,还不如说是"人工智能的合规要求";更重要的是,法案文本洋洋洒洒,立法背景和法律原则阐述的十分充分,但对这些条款的具体执行语焉不详。比如,第 16 条"提出对操纵性和剥削性实践的禁止",但如何"禁止操纵性和剥削性实践"仍需具体的实施性规则予以落实。该法案第 81 条中还采用了"鼓励"的方式,授权提供者自主制定行为规则,以使非高风险人工智能系统符合"合乎道德和值得信赖"的标准。

我国前几年颁布的涉数据和算法监管的相关法律法规和行政规章中,虽对人工智能伦理的相关原则和标准有所涉及,表述却也都较为间接。比如《数据安全法》第 8 条、第 28 条,只提及要"遵守社会公德和伦理",但对于什么是社会公德、什么是伦理,为什么要区分道德和伦理,相关条款及公开的立法说明也语焉不详。国内近来公开的两个《人工智能法(专家建议稿)》,〔12〕无例外地都对人工智能伦理原则及其审查制度做了专门的条款拟制,但相关内容仍然抽象。全国网络安全标准化技术委员会于 2024 年 3 月发布《生成式人工智能服务安全基本要求》等类似的技术标准可以理解为是人工智能伦理合规可操作化的一种形式,但对于大部分的合规主体来说,它在伦理要求的标准化方面仍然是间接且不明确的。

那么该如何将抽象、原则的人工智能伦理规范转变为可操作的伦理合规实践呢?初步概括,大致上可以分为三个步骤:一是"(人工智能伦理)是什么"的确定,主要的操作模式是将人工智能伦理的实质性内容概括为一系列原则。比如"可信、安全和负责任 AI"的提法。二是"人工智能伦理如何实施"的操作,即要将人工智能伦理原则细化为相关的规则和标准。比如,

〔8〕同上注;see F. Doshi-Velez, M. Kortz, R. Budish et al., *Accountability of AI under the Law: The Role of Explanation*, Social Science Electronic Publishing, 2017, p. 7.

〔9〕施敏、杨海军:"生成式人工智能的算法伦理难点分析与探索",《大数据》2024 年第 2 期,第 9 页。See David Oniani, Jordan Hilsman, Yifan Peng et al., "Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare," *npj Digital Medicine*. Vol. 6, No. 1, 2023, p. 225.

〔10〕参见吴红、杜严勇:"人工智能伦理治理:从原则到行动",《自然辩证法研究》2021 年第 4 期,第 49—54 页;Ramak Molavi Vasse'i, "The Ethical Guidelines for Trustworthy AI-A Procrastination of Effective Law Enforcement," *Computer Law Review International*, Vol. 20, No. 5, 2019, pp. 129-136.

〔11〕Ibid.

〔12〕参见中国社科院法学所周辉课题组起草的《人工智能示范法》和中国政法大学张凌寒课题组起草的《人工智能法专家建议稿》。

将可信人工智能的原则标准化为相关产品性能的合格率、将负责任人工智能的原则分解为风险责任分担配置规则等。在这个阶段，除了上述实体性规则，还涉及专家委员会职权配置以及程序设计。三是“人工智能伦理实施状况评估”，包括将规则和标准应用到实际产品研发和应用中，以验证特定人工智能系统、服务或者产品是否符合相应的伦理原则。〔13〕在这里，真正的技术难点在于第二步，同时也是法学介入人工智能伦理讨论的关键点。我们可以将其转化为人工智能伦理的法化过程来展开讨论。需注意的是，不能过分夸大法化、“法治化”思路的重要性，而忽视技术方法在此阶段的重要性。比如，价值对齐测试既是一个技术性问题，也是牵涉人文社会科学整体知识的过程。

(二)法化思考路径的基本考虑

法化(Verrechtlichung，juridification)是一个多义且富有批判性内涵、兼有描述性和规范性的概念。一般人认为，法化包括“构成性法化”、法律的扩张和分化、通过正式法律途径解决冲突的增长、司法权力的增长、法律文化的扩张等。〔14〕

从马克斯·韦伯到卢曼，从欧洲的社会理论到美国的法律与社会运动，再到日本世纪之交关于“法制现代化与过度法化”的讨论，法社会学、法理学领域对法化问题做了较多探讨。〔15〕比如，田中成明将法化概括为三个面相：法的要求的增强、法规范或者制度的复杂化，以及法律价值、原理、规范和程序在人们的意识和行为中的内化。〔16〕在此分类基础上，田中成明和同时期的日本学者，比如棚濑孝雄、六木佳平等学者一起，在美国批判法学的理论基础上，认定“法治主义”在日本二战之后的实践产生了异化现象。〔17〕他通过一种掺杂着规范追求和事实描述的类型化思路，将这种因过度法化的法治异化的出路概括为两种：一是，“非法化”或者“反法化”，其中包括法律程序的非正式化，比如 ADR 等非正式纠纷解决机制的兴起；二是法的范畴和种类的扩展，即出现了管理型法和自治型法。〔18〕

随着前沿科技的兴起，近来国内学术界也有学者，比如朱芒试图通过法化的概念来研究技术标准、伦理规范、学校内部规范等新型“制度形态”。〔19〕在此，延续此种研究思路，试图将法化的概念导入人工智能伦理的研究过程。为理解方便以及在中国语境下展开分析的需要，此

〔13〕 See Jessica Morley, Luciano Floridi1, Libby Kinsey et al., “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices,” *Science and Engineering Ethics*, Vol. 26, No. 4, 2020, pp. 2141-2168.

〔14〕 See Penny Brooker, “The Juridification of Alternative Dispute Resolution,” *Anglo-American Law Review*, Vol. 28, No. 1, 1999, pp. 1-36.

〔15〕 已有的中文文本中，较为集中的讨论可以参见(日)田中成明：《现代社会与审判：民事诉讼的地位和作用》，郝振江译，北京大学出版社 2016 年版。

〔16〕 同上注，第 26 页；長尾龍一＝田中成明『現代の法哲学』(東京大学出版会，1983 年)76—77 頁参照。

〔17〕 参见(日)六木佳平：《日本法和社会》，刘银良译，中国政法大学出版社 2006 年版；(日)棚濑孝雄：《现代日本的法和秩序》，易平译，中国政法大学出版社 2022 年版。

〔18〕 参见田中成明，见前注〔15〕，第 28 页。

〔19〕 参见朱芒：“行政规范性文件的功能结构”，《法学家》2023 年第 6 期，第 57—71 页。

处将法化简化定义为:社会道德、宗教伦理、职业道德、技术标准、行业习惯等社会规范被法律化的过程或者现象。

需要将法化与两个概念进行辨析:一是与法律渊源的区别。“法化”与“法律渊源”两个概念的主要差异在于,前者是一个过程性描述;它不仅可以用来描述“社会规范”被转化为“法律渊源”的过程,而且还可以用来描述权力形态的变迁(比如司法权的扩张)、法律意识和文化的变化,等等。此外,法化还会牵涉权利化、程序化等内涵。二是与中国学术界和实务界惯用的“法治化”的区别。在人工智能伦理与法律实践关联的学术研讨中,也有很多学者习惯于运用“法治化”“规范化”等动态概念。〔20〕 此处重点解释一下,为什么使用“法化”而放弃“法治化”概念的主要原因:

第一,国内学术界对“法治化”概念使用的泛化和概念内涵的窄化。从“法治”之中蕴含的法律至上、权利本位、正当程序、权力制约以及良法之治等原则,我们可以演绎出,在“法治化”的技术操作中,包含着法律发现、权利化、利益衡量、程序保障等多重内容。但由于种种原因,在当下我国,“法治化”的概念有被不断泛化使用但内涵却不断窄化为“制度化”(即“法制”,rule by law,区别于 institutionalization 意义上的“制度化”)的趋势。

第二,“法治化”内涵中缺乏批判性的面相。在人工智能伦理及其立法的过程中,在国内外都存在着人工智能伦理泛化以及道德和法律混淆、在法律合规之外增设人工智能伦理审查制度的做法提出了批评。〔21〕 相关质疑从欧盟的数据立法开始一直持续到人工智能立法过程。这种对道德泛化的批判也可一直上溯至生物医药伦理大行其道的 20 世纪七八十年代。在当时,各种社会理论、批判性法学理论对“法化”“权利话语”的讨论,也多会涉及这一点。〔22〕 简言之,法治化是一个正面概念,法化的概念则正如世纪之交的日本学者所讨论的,它更容易让我们观察到“法治异化”的侧面;在应对法治异化的制度措施上,更能够提出“去法化”“反法化”等更加多维、多方向的理论概念和解决措施。〔23〕

二、人工智能伦理如何法化:通过事物本质的探寻

(一)伦理、技术与法律的共生重构

通过回顾各国政府、科研院校以及产业界发布的各种原则、准则、指南、框架以及清单,人

〔20〕 参见赵鹏:“生物医学研究伦理规制的法治化”,《中国法学》2021 年第 2 期,第 25—44 页;宋华琳,见前注〔7〕,第 1—9 页。

〔21〕 Ben Wagner, “Ethics as an Escape from Regulation: From ‘Ethics-Washing’ to Ethics-Shopping?” in Mireille Hildebrandt (ed.), *Being Profiling: Cogitas ergo sum*, Amsterdam: Amsterdam University Press, 2018, pp. 84-89; Paula Boddington, *Towards a Code of Ethics for Artificial Intelligence*, Oxford: Springer, 2017.

〔22〕 See Steve J. Bickley and Benno Torgler, “Cognitive Architectures for Artificial Intelligence Ethics,” *AI & Society*, Vol. 38, No. 2, 2023, pp. 501-519.

〔23〕 参见朱芒,见前注〔19〕。

工智能伦理的实体内容绝大部分被已有的网络、数据、算法以及个人信息保护等方面的法律原则或者规则所涵盖。我们可以延续科技伦理法化初期的讨论，提出一个疑问：既然已有的立法文本基本涵盖了这些道德原则或者伦理原则，那为什么在技术治理中，一定要将伦理和法律并列？法律与道德区别的传统标准之一在于：国家强制力。但人工智能伦理审查制度的法律效力阐述是：没有建立伦理审查制度或者没有通过具体伦理审查程序的某些特定人工智能系统不能开展研发或者无法进入市场。在此背景下，通过转致条款、授权条款等，作为道德内容的人工智能伦理实体性规则，即使不会被实体法明确表述，但也具有了类似于标准等规范性文件的国家强制力。〔24〕那么，在学理上以及实务操作中该如何把握两者的异同呢？

提倡人工智能伦理制度独特性的监管者和学者的大致回答是：在功能上，法律合规还不足以引导社会朝着正确的方向发展。因为法律监管的两元代码是合法和非法，它还无法说明在合法行为中，什么才是好的或者最好的行动，或者符合善治的要求。〔25〕另外，职业伦理、商业伦理以及技术伦理等应用伦理与法律之间的耦合、互动演化关系的成熟讨论，可以作为应对上述质疑的答案。〔26〕顺着这种思路，也可以说，科技伦理的法化其实也是“科技治理的伦理化”的重要内容；随着科学技术对社会结构带来的深远影响、对人类道德生活和利益分配带来的深刻影响，法律、伦理以及技术三者之间，出现了深度的嵌合重构。换言之，它深刻地嵌入到了社会、经济、政治和文化结构，对人类道德生活和利益分配的冲击日益深远，国家不得不通过法律手段介入并将伦理与国家法律和监管体系相嵌合。〔27〕为了更好地进入正题讨论，此处暂时按下技术道德法化必要性讨论不提，而转向“如何法化”“如何更好地法化”等环节，为人工智能伦理的合规实务提供理论指引。

（二）科技伦理法化的已有思路：新兴权利、软法与共同体伦理

已有关于科技伦理法化的研究，主要存在三种思路：

一是，道义论的应对思路：人格权等新兴权利研究。有私法学方向的学者从人格权保护，提倡从立法上对科技伦理进行规范。这方面国内的法学者代表有中国人民大学的石佳友等人。〔28〕但类似的新兴权利探讨，并不只局限于私法学者。比如，针对神经技术、脑机接口技术引发的伦理困境提出的神经权利研究，主要是一批神经技术专家和伦理学家，且主要在

〔24〕 参见朱芒，见前注〔19〕。

〔25〕 See Boddington, supra note 21；赵鹏：“科技治理‘伦理化’的法律意涵”，《中外法学》2022 年第 5 期，第 1201—1220 页。

〔26〕 Gunther Teubner and Altera Pars Audiatur, “Law in the Collision of Discourses,” in Richard Rawlings(ed.), *Law, Society and Economy: Centenary Essays for the London School of Economics and Political Science*, Oxford: Clarendon Press, 1997, pp. 1895-1995. 中文简介的文献综述可以参见刘闯：“伦理与演化博弈：道德起源与本质的哲学探究”，《哲学研究》2024 年第 2 期，第 91—101 页。

〔27〕 参见赵鹏，见前注〔25〕。

〔28〕 参见石佳友、刘忠炫：“人体基因编辑的多维度治理——以〈民法典〉第 1009 条的解释为出发点”，《中国应用法学》2021 年第 1 期，第 171—188 页；石佳友、徐靖仪：“医疗人工智能应用的法律挑战及其治理”，《西北大学学报（哲学社会科学版）》2024 年第 2 期，第 91—103 页。

宪法学、国际公法学的思路进行推进。〔29〕有一些思路则从权利束的角度试图突破人格权保护不足的问题——当然后者的思考基于一种私法意义上人格权保护不足的问题意识。〔30〕此外,还有学者实质上提出了“权利动态化”以应对科技发展带来的伦理冲突困境。〔31〕

二是,结果论的应对思路:软法研究。在国内外法学界,公法学者会从软法和硬法的两元分立框架来讨论科技伦理的功能及其法律性质,国内的学者代表有北京大学的沈岿教授。〔32〕国际公法学者对此也有类似的体系阐述。〔33〕中国政法大学的赵鹏关于科技伦理法治化的研究、〔34〕南开大学法学院宋华琳关于人工智能伦理法治化、〔35〕上海政法学院刘长秋关于生命伦理法律化〔36〕的思路,都可归入这一研究进路。其中,赵鹏的重要法化思路是:通过科技与社会关系的不确定性分析,提出部门法介入科技伦理治理的必要性及其优势。在他们的研究中,对法化过程的描述,主要采用更加正面的“法治化”的概念。在英文世界,除了法学界之外,公共管理学界、伦理学界以及经济学界都有充足的研究成果。〔37〕软法的研究思路明显受到美国规制学派或者新行政法学的影响,具有鲜明结果论或者功利主义的理论传统,与公共管理学、社会学等学科中的人工智能治理研究形成了交叉关系。

三是,美德论的应对思路:共同体伦理的思路。持这一研究思路的学者主要是一些来自管理学、科学学、应用伦理学等专业的学者。比如在关于如何将人工智能伦理原则转变为合规实践中,有较多的应用伦理学以及人工智能技术专家将希望寄托在美德伦理学思路。这种思路的主要特点是:不定义人工智能伦理行为准则的具体内容,而是关注技术开发主体的个人层面,特别是技术专家和工程师的社会背景;通过增强技术人员及其所在企业的社会责任感、追求美德目标等来消除“技术中立性”原则的弊端。〔38〕这实质上就是美德伦理学思路上的“(工程师)职业伦理”和“商业伦理”。当然也存在康德哲学、道义论意义上的工程师伦理思路:事前

〔29〕 参见李学尧:“‘元宇宙’时代的神经技术与神经权利”,《东方法学》2023 年第 11 期,第 74—84 页。

〔30〕 参见王锡锌:“重思个人信息权利束的保障机制:行政监管还是民事诉讼”,《法学研究》2022 年第 5 期,第 3—18 页。

〔31〕 参见王贵松:“风险行政与基本权利的动态保护”,《法商研究》2021 年第 4 期,第 18—31 页。

〔32〕 参见沈岿:“软法助推:意义、局限与规范”,《比较法研究》2024 年第 1 期,第 148—164 页。

〔33〕 参见朱明婷、徐崇利:“人工智能伦理的国际软法之治:现状、挑战与对策”,《中国科学院院刊》2023 年第 7 期,第 1037—1049 页。

〔34〕 参见赵鹏,见前注〔20〕;赵鹏,见前注〔25〕。

〔35〕 宋华琳:“法治视野下的人工智能伦理规范建构“,《数字法治》2023 年第 6 期,第 1—9 页。

〔36〕 刘长秋:“生命伦理法律化研究”,《浙江学刊》2008 年第 3 期,第 141—146 页;Thilo Hagendorff, “The Ethics of AI Ethics: An Evaluation of Guidelines,” *Minds and Machines*, Vol. 30, No. 1, 2019, pp. 99-120.

〔37〕 Ryan Calo, “Artificial Intelligence and the Carousel of Soft Law,” *IEEE Transactions on Technology & Society*, Vol. 2, No. 4, 2021, pp. 171-174; Johan Rochel, “Learning from the Ethics of AI: A Research Proposal on Soft Law and Ethics of AI,” *Tilburg Law Review*, Vol. 27, No. 1, 2022, pp. 37-59.

〔38〕 See Rochel, ibid.; 中文的类似思路,参见孟令宇等,见前注〔2〕。

确定一套固定的通用性原则和规则，让技术开发主体去严格遵守。〔39〕

上述研究思路在实践中时常是融合在一起的。比如，以生物医学伦理为主的科技伦理的监管思路是：在研发阶段，将其建构为科学家共同体的职业伦理，在技术研发阶段通过“名誉机制”和“逐出职业群体的威胁”为主的预防性、自我规制机制来实现风险治理的目的。〔40〕而在社会应用阶段，科技伦理的重点则转移到相关利益方的充分认知和有效评估，在这种思路中，作为职业伦理自我规制规则体系的生物医学伦理，通过各类立法将生物医学伦理规则嵌入到公法意义上的行政许可前置条件、违法或者违反行政秩序行为的确认、非法行为的问责（包括侵权责任配置甚至刑罚对象）等兼有预防性和修正式的方式，不断地提升其法律效力，从而逐渐演变成了兼有“软法”和“硬法”、跨越公私法界限、超越国家法和社会规范两分的规范体系。〔41〕

（三）从职业伦理到科技伦理：作为参照系的生物医药伦理

生物医药伦理（bioethics）最早可以上溯至著名的“希波克拉底誓言”。该誓言呼吁医生要帮助患者（仁慈）、不伤害（非恶意），并遵守保密义务。此后，在希波克拉底誓言基础上，伴随着近现代“权利社会”、福利社会的到来，生物医药伦理还生长出了以患者自主权为核心的患者利益中心主义（自治）、不能因患者个人和个体特征而受到歧视（反歧视）等原则。〔42〕从性质定义，医生伦理在此阶段的性质属于职业伦理（professional ethics），且是比教师伦理、法律职业伦理形成更早、更模范性的一种职业伦理。

随着现代科学研究和技术探索，不断地从纯思辨的理论知识活动，演变成一种有目的、呈现规模化的实际行为，特别是科学成果的发现方式与途经对社会以及科研活动过程中所涉及到的人群或者人类所珍惜的某种价值（比如“美好环境”）有着某种危害，甚至会引发社会风险，即科技活动在很大程度上不再是一种价值中立的行动，于是在现代西方出现了技术伦理、工程伦理等科技伦理的讨论，并最终聚合产生科技伦理的概念。〔43〕作为一种责任伦理，科技伦理在早期实质上也是一种职业伦理意义上的“科学家伦理”。在英文世界，很多关于技术伦理、工程伦理的教材或者通识论著，往往直接将上述伦理定义为职业伦理。

在此阶段，生物医药伦理逐渐从聚焦于医疗实践的医生伦理逐渐演变成更广泛地思考科

〔39〕 See Brent Mittelstadt, “Principles Alone Cannot Guarantee Ethical AI,” *Nature Machine Intelligence*, Vol. 1, No. 11, 2019, pp. 501-507; Brent Mittelstadt, Chris Russell and Sandra Wachter, “Explaining explanations in AI,” in *FAT * '19: Proceedings of Conference on Fairness, Accountability, and Transparency*, January 29-31, New York: Association for Computing Machinery, 2019, pp. 279-288.

〔40〕 See Boddington, supra note 21, p. 54. 具体机制的阐述，参见赵鹏，见前注〔25〕。

〔41〕 See Lorne Sossin and Charles Smith, “Hard Choices and Soft Law: Ethical Codes, Policy Guidelines and the Role of the Courts in Regulating Government,” *Alberta Law Review*, Vol. 40, No. 4, 2003, pp. 867-893. 中文文献可参见赵鹏，见前注〔20〕。

〔42〕 See Dennis A. Robbins, Frederick A. Curro and Chester H. Fox, “Defining Patient-Centricity: Opportunities, Challenges, and Implications for Clinical Care and Research,” *Therapeutic Innovation & Regulatory Science*, Vol. 47, No. 3, 2013, pp. 349-355.

〔43〕 参见甘绍平：“科技伦理：一个有争议的课题”，《哲学动态》2000年第10期，第5—8页。

学目的、生命本质等问题的“生命伦理”。〔44〕伴随着克隆技术、基因编辑等技术的进展，相比于核科学、工程建设等通过国家强监管机制即得以实现安全目标的科学研究和技术探索，出于对自身随时可能会成为“天竺鼠”（实验小白鼠）的恐惧，在全球意义上，生命伦理议题在近三四十年成为学术界和社会层面的关注焦点。

由于科技活动性质的变化，此阶段的科技伦理逐渐出现与职业伦理相脱离的趋势。因为科学技术的发展动力源于现实需求，开发什么样的技术、应用于什么样的领域以及如何应用这些技术，已不是由科学家或者科技从业者决定，而是取决于掌握科学技术权力的人们，〔45〕包括公权力和资本。因此，责任伦理不断扩展到科学家或者科技从业者之外的其他人员，比如“实验室的管理单位及其负责人”；伦理合规的认定方式也逐渐从“职业共同体认定”演变为“有法律专家、伦理专家等非职业共同体内部人员共同参加”的“伦理审查”。

本处简要回溯了生物医药伦理的性质变迁史，即它从医生的职业伦理，到科学家的职业伦理，再到“超越职业伦理性质”的科技伦理的演变。这种回溯并不是为了研究生物医药伦理的历史，而只是试图通过生物医药伦理性质变迁从而导致其规范内容变化历史的回顾，将人工智能伦理的发展续接在这一传统中，从而有效探知其规范内容。

（四）到法学方法论寻求理论工具：事物本质

人工智能伦理的法化，既要以已经法化且理论论辩充分的生物医药伦理为参照，〔46〕也要依据人工智能的技术特征对其法化进行实践操作，具体步骤有二：一是关注并概括人工智能伦理的具体特征，这些特征对于提高人工智能设计和开发的道德可持续性具有重要意义；二是采用有助于解释人工智能伦理相关特征的方法论工具，这一方法论工具应能有效处理、架接事实与价值之间的鸿沟。〔47〕那么，法学研究者在过去是如何尝试凿穿事实与价值的壁垒（或架接鸿沟）的方案？我们很自然会想到要去法学方法论的工具箱里寻找理论工具。

在这里，我们试图导入在国内法学界逐渐归于沉寂的适合于推论的“事物本质”这一理论概念。尽管有着诸多的定义，以考夫曼的理论为主，“事物本质”的方法包含以下三个命题：①事物拥有本质，并外显为不以人的意志为转移的特征或者规律；事物的“本质”介于“建构性理念—规范”与“事物”（法律所面对的具体生活关系）的中间。〔48〕②“事物本质”，既有来自普世永恒的属性，也有来自不同的具体生活关系的特殊的、历史的因素，因此它具有规范性意义。③“事物本质”可能会因实然而获得规范效力，也就是说，“事物本质”具有法律渊源的特性。从这个角度来说，法律规范的结构和内容应当附随于事物本质的规范性意义。〔49〕据此，事物本质的理论要点可概括为“法理念或法律规范必须与生活事实维持一致，彼此应相互适应”。

可以举例说明事物本质理论应用的典型性场景。比如，2012 年《民事诉讼法》第 112 条将

〔44〕 参见赵鹏，见前注〔20〕。

〔45〕 参见庄友刚：“风险社会中的科技伦理：问题与出路”，《自然辩证法研究》2005 年第 6 期，第 71—75 页。

〔46〕 比如，在考夫曼的《法律哲学》一书中，甚至将“生命伦理学”等作为法律哲学的重要相关内容做了重点阐述。参见（德）阿图尔·考夫曼：《法律哲学》，刘幸义等译，法律出版社 2011 年版，第 408—476 页。

〔47〕 See Bickley and Torgler, supra note 22.

〔48〕 （德）阿图尔·考夫曼：《类推与“事物本质”》，吴从周译，学林文化事业出版社 1999 年版，第 21 页。

〔49〕 （德）阿图尔·考夫曼：《当代法哲学和法律理论导论》，郑永流译，法律出版社 2013 年版，第 125 页。

“民事虚假诉讼”界定为“当事人之间恶意串通，企图通过诉讼、调解等方式侵害他人权益”的诉讼行为，但实际上，“虚假诉讼”的本质是当事人虚构案件纠纷以实现侵害他人合法权益和社会公共利益的目的，既可以是主体间串通，也可以不经串通而由当事人单方捏造证据虚构法律关系来实现，所以，一旦明了虚假诉讼依其本质表现出的特征和规律，就可明确当事人之间串通和一方当事人捏造这两种情形均体现虚假诉讼的社会危害性，故均需要通过法律予以规制，以此类推形成民事虚假诉讼法律规范的“恰当内容”。〔50〕

可以采取与法化概念类似的思路，将事物本质的理念纳入以嫁接“人工智能的特征”与“人工智能伦理的特征”，即为了有效法化，广义的立法者或者“找法者”从人工智能技术及其产业的特征中推导出（规范性的）伦理要求的过程。〔51〕一言遮之，此处关于“事物本质”的应用场景，从法律适用场域转移到了广义的立法情景。

学术界对“事物本质”理论的批判，主要集中于：①在前提上，抹杀了实然和应然的区分。②在方法上，具有非科学性。与其相关的类推、类型等概念具有危险的不确定性，容易沦为恣意的工具。〔52〕在人工智能伦理的法化过程中，运用“事物本质”这一概念展开分析，其优势至少可以表现在两个方面：①在立法的场景里，因立法的民主化过程（商谈、利益博弈）的性质所决定，相比法律适用场景，它对于法律方法论科学性的要求要大大降低，从而能够有效弱化前文提及的反对事物本质相关理论的批判力，并在法律实践中较为可行地激活这一理论。②借助于推论与类型化等方法，能够更好地参照已有被有效法化且理论论辩充分的科技伦理（技术伦理）——生物医药伦理，进而能够从人工智能技术特征及其引发的社会关系变革入手来提出规范性规范要求的方法。

三、人工智能伦理法化所依赖的事物本质：三大背景性约束

在上述思路指引下，大致可以概括出人工智能伦理法化过程无法回避的三个特征或者约束性条件：具有道德规则的技术可嵌入性、更强的场景性以及依赖于技术过程的程序性。

（一）道德化人工智能与技术可嵌入性特征

人工智能伦理的可嵌入性主要基于“道德化人工智能”的研发思路，即为了实现伦理原则的可操作性，以及考虑人工智能应用的场景化特征，从人工智能产品的开发最初，就要将嵌入式伦理整合到算法中，从而达到预测、识别并解决人工智能开发和应用过程中的伦理问题。〔53〕道德化人工智能（ethical AI）主要是指遵循透明、公平、责任、隐私保护等人工智能伦理的算法、架构和接口。这种人工智能概念的提出，实质上就是基于人工智能伦理实践化的重要思路：让抽

〔50〕 参见王亚新：“虚假诉讼的法律规制——特集导读”，《交大法学》2017年第2期，第5—6页。

〔51〕 参见陈爱娥：“事物本质在行政法上之应用”，《中国法律评论》2019年第3期，第82—92页。

〔52〕 参见（德）魏德士：《法理学》，丁晓春、吴越译，法律出版社2013年版，第204页。文献综述可参见姜纪超：“‘事物本质’及其法学方法论意义”，载《法律方法》（第8卷），山东人民出版社2009年版，第437—448页。

〔53〕 See Edmond Awad, Sohan Dsouza and Richard Kim et al., “The Moral Machine Experiment,” *Nature*, Vol. 563, No. 7729, 2018, pp. 59-64.

象的伦理原则变成可操作性的现实。〔54〕与之相关，人工智能的嵌入性特征主要是指，在产品研发或者设计中，通过深度集成、协作和跨学科的方式，直接将人类普遍认可的伦理原则编程到算法中，以使其“合乎伦理”，并让其自主或者半自主地自我演化。〔55〕

人工智能伦理的可嵌入化研究并不新鲜。早在 20 世纪中叶，阿西莫夫的机器人三定律就是伦理可嵌入性的理想化表述。这种思路的主要内容就是：通过技术来确立并实现机器人的道德法则体系。用科学哲学的话语来表述就是：采取自然主义方法论来嵌入功能性道德。它的构建策略基本有三：其一是自上而下，即在智能体中预设一套可操作的伦理规范，如自动驾驶汽车应将撞车对他人造成的伤害降到最低。其二是自下而上，即让智能体运用反向强化学习等机器学习技术研究人类相关现实和模拟场景中的行为，使其树立与人类相似的价值观并付诸行动，如让自动驾驶汽车研究人类的驾驶行为。其三是人机交互，即让智能体用自然语言解释其决策，使人类能把握其复杂的逻辑并及时纠正其中可能存在的问题。〔56〕按照科学哲学主流的阐述，通过这些途经，可以将人工智能纳入道德共同体，进而通过三种方式影响人机交互：其一，人工智能将和动物一样得到人类的伦理 关注；其二，人类将认为可以对人工智能的行动进行道德评估；其三，人类将把人工智能作为道德决策的论证和说服目标。但这些策略都有其显见的困难：如何在代码化和计算中准确和不走样地表达与定义伦理范畴？如何使智能体准确地理解自然语言并与人进行深度沟通？〔57〕

人工智能伦理嵌入性研发，在实践中，公开可见的存在于涉生物医药的人工智能产品研发中。比如，在哈佛大学医学院乔治·邱奇(George Church)的合成生物学实验室里就有专职的伦理学家；而德国慕尼黑工业大学的机器人和机器智能学院(MSRM)在生产人工智能驱动的医疗产品时，就聘请了慕尼黑医学历史与伦理研究所等单位的伦理学家和法律专家。〔58〕此外，在笔者参与的一些智慧司法系统研发中，某些将法律程序的算法程序化工作，实质上也是这种思路的重要组成部分。〔59〕尽管人工智能伦理的嵌入性研发仍然还在路上，但是，这完全可以确立这一伦理规则的嵌入性特征，并提示我们有必要将伦理监管的思路落实到技术研发的嵌入式过程。

---

〔54〕 See Barry Dewitt, Baruch Fischhoff and Nils-Eric Sahlin, “‘Moral Machine’ Experiment is no Basis for Policymaking,” *Nature*, Vol. 567, No. 7747, 2019, p. 31.

〔55〕 See Louise Bezuidenhout and Emanuele Ratti, “What does it Mean to Embed Ethics in Data Science? An Integrative Approach Based on Microethics and Virtues,” *AI & Society*, Vol. 36, No. 3, 2020, pp. 1-15.

〔56〕 法学意义上从人机交互的角度对伦理嵌入做深入阐述的系统阐述可以参见季卫东、赵泽睿：“人工智能伦理的程序保障——法律与代码的双重正当化机制设计”，《数字法治》2023 年第 1 期，第 57—76 页。

〔57〕 段伟文：“开放的伦理底线与结构化伦理嵌入——深度科技化时代的生命伦理审度”，《探索与争鸣》2018 年第 5 期，第 15—16 页。

〔58〕 Nora von Ingersleben-Seip, “Competition and Cooperation in Artificial Intelligence Standard Setting: Explaining Emergent Patterns,” *Review of Policy Research*, Vol. 40, No. 5, 2023, pp. 781-810.

〔59〕 智慧司法研发中涉及的自动化决策引发的法理学探讨，可以参见徐舒浩：“自动化决策系统语境中不作为因果关系之司法证明”，《地方立法研究》2024 年第 2 期，第 69—90 页。

(二)信息高频流动与更强的场景性

人工智能技术引发的道德挑战以及伦理回应的思路,与传统生物医学伦理原则非常相似,甚至可以与其共享绝大部分的基本原则,特别是"增进福祉"与"风险的合理控制"。但是,与信息技术兴起之前的生物医学伦理不同的是,人工智能领域的伦理问题具有更强的不确定性、发展性,从工业应用到自动驾驶,从养老到医疗,再到法律科技,场景的复杂性令人眼花缭乱,在场景化的伦理审查或者评估过程中,同时需要制定适应于该应用场景的相关伦理规则。相比传统的生物医药技术,人工智能技术(包括脑机接口等获人工智能助力的生物医药技术)的迭代速度更快,容不得技术开发者在较长的时间去展开相关的伦理冲突评估。也正是如此,有专家认为,人工智能伦理产生作用的最大障碍是可操作问题。〔60〕

海伦·尼森鲍姆(Helen Nissenbaum)针对信息技术带来的隐私权保护难题,特别是个人信息难以界定、知情同意原则式微与虚化、信息处理参与者利益失衡等诸多挑战,提出了隐私场景公正理论(contextual integrity theory,以下简称场景理论)。该理论的核心观点是,应结合具体场景进行针对性保护,提倡具体的风险防控、反对泛化的个人信息保护;提倡采用信息处理的动态思路、反对个人信息固定不变;提倡采用宽容和促进产业发展的监管思路、反对个人主义的理念。〔61〕这种场景理论,尽管是从社会学意义上叙事理论展开阐述的,但是也说明了人工智能技术以及其伦理规则高度场景化的特征。

与传统生物医学伦理相比,人工智能伦理(包括将人工智能系统应用于传统生物医学领域而引发的伦理领域)在场景性方面表现为以下几个方面:

一是个人信息界定的难题。在一种"鼓励数据流通"的思路下,信息或者数据流动的链条无限被拉长、隐私和信息之间的界限越来越模糊,〔62〕信息脱离原来的适用场景之后可能会被无数次地重新使用,最初使用的合理性基础早已不存在,而在每次重新使用过程中,该信息是否属于"需要保护的个人信息"以及如何"分类分级地进行保护",都取决于具体场景的性质、场景中各方主体利益的均衡以及比例原则等多种相互叠加的法律、伦理原则和规则的适用权衡。〔63〕当然,在传统生物医学伦理领域也有类似问题,比如在1951年的海拉细胞系(Hela Cell Line)案件中,从海里埃塔·拉克丝(Henrietta Lacks)身上提取的癌症细胞被反复复制并被生物医药公司用以牟利,就是一种场景转换带来的科技伦理问题。〔64〕

二是关于知情同意权的虚化问题。在前信息技术时代,通过生物医药技术研发实践逐步

---

〔60〕 Jianlong Zhou and Fang Chen, "AI Ethics: From Principles to Practice," *AI & Society*, Vol. 38, No. 6, 2023, pp. 2693-2703.

〔61〕 See Helen Nissenbaum, "Privacy as Contextual Integrity", *Washington Law Review*, Vol. 79, No. 1, 2004, pp. 119-157.

〔62〕 参见王锡锌,见前注〔30〕。

〔63〕 参见高莉:"大数据伦理与权利语境——美国数据保护论争的启示",《江海学刊》2018年第6期,第151—156页;王锡锌:"国家保护视野中的个人信息权利束",《中国社会科学》2021年第11期,第115—134页。

〔64〕 See Diana-Lyn Baptiste, Nicole Caviness-Ashe and Nia Josiah et al., "Henrietta Lacks: Science Must Right a Historical Wrong," *Nature*, Vol. 585, No. 7823, 2020, p. 7.

形成的知情同意制度，被认为是个人信息保护制度的关键内容。但随着信息时代的到来，这一制度的虚设性越来越明显。相关领域学界已经有了较为丰富的研究成果，此处不再赘述。

三是风险责任的分担问题。顺着对主流法学影响甚深的伦理学或者道义论的思路，前沿科技引发的风险责任自然应由“作为技术开发者或者服务提供者的加害人”承担。如果严格遵循这种“民间心理学”或者“直觉道德感”的思路，在法律适用中，就不大可能会出现保障美国和中国网络平台经济发展的“避风港原则”和“红旗规则”。但是，令人忧虑的是，在短短的二十多年中，伴随着数字技术的不断迭代，“避风港原则”在中美的监管实践中都已经出现多次实质性的修改。〔65〕 此外，关于算法其实、社会公平、风险控制等，全球背景下，对相关伦理问题的认识也伴随着技术的发展而不断迭代，令人难以适从。

在学术界，在场景化特征方面，关于人工智能伦理和生物医药伦理的区别中，还涉及能否“职业伦理化”的讨论。有学者认为，因为生物医药技术的“封闭性”特征，使得生物医药伦理是可以转化为高等学校教师、科研人员的“职业伦理”，但是人工智能几乎适用于任何人类，因此无法产生一个“使用人工智能”的职业或者工种伦理。这一讨论可能涉及对人工智能伦理内容的分类分级。〔66〕

（三）透明化、可解释性难题与程序性特征

在人工智能伦理中，透明度、可解释以及可问责是一组相邻、链条化的原则。关于这些伦理原则的技术缘起以及技术化思路已经有很多文献展开讨论。〔67〕 在生成式人工智能出现之前，为了应对算法歧视等影响社会公平的问题，以算法透明、可解释为内核，各国逐渐构建起一套完整的算法治理体系。但是，由于基于大模型的生成式人工智能复杂的内部工作机制，使得其具有透明度低、可解释难、算法归责存在困境等问题。尽管学术界普遍认为，不可能解释的“涌现”现象，只是暂时不能解释而已，但它对此前的算法治理体系形成了严重的冲击，甚至颠覆。《暂行办法》试图通过“从算法治理到模型治理”的治理模式革新，解决这个问题，但是在可执行性方面仍然还有很远的路要走。

在此背景下，理想主义地从完全的可解释性来给开发者或者应用者配置责任，显得十分不现实。〔68〕 这时候我们或许可以借鉴“职业伦理是一种程序伦理”的思路，回到正当程序的思路来进行规制人工智能伦理。当然，运用正当程序原则来规制人工智能的应用，特别是对公权力的限制并不是一个新话题。〔69〕 但是如何贯彻技术性正当性程序原则，仍需落实为技术实

〔65〕 参见赵泽睿：“平台革命引发的美国版权责任变革及经验分析”，《电子知识产权》2020 年第 12 期，第 34—48 页。

〔66〕 See Paula Boddington, “Towards a Code of Ethics for Artificial Intelligence,” in Barry O’Sullivan and Michael Wooldridge (eds.), *Artificial Intelligence: Foundations, Theory, and Algorithms*, Berlin: Springer, 2017, p. 93.

〔67〕 See Bickley and Torgler, supra note 22, pp. 501-519.

〔68〕 See Elizabeth A. Holm, “In defense of the Black Box,” *Science*, Vol. 364, No. 6435, 2019, pp. 26-27.

〔69〕 季卫东：“主权的嬗变——数字化‘魔兽世界’与法律秩序创新”，《交大法学》2023 年第 5 期，第 5—17 页。

践问题，这就需要回到前文所谈及的伦理技术嵌入性的问题。我们可能需要更宽广的视野，从人工智能认知架构设计与程序机制内置化的角度进行展开。

对于以往的生物医药伦理而言，程序法和实体法的性质争论实质上是不存在的：生物医药伦理的程序性伦理，主要落实在“审查”，而此前的生物医药伦理的实体内容是相对确定的。但是，人工智能伦理强烈的场景化特征，使得其很多伦理规则需要在伦理审查现场产生，在此背景下人工智能伦理的过程性、程序性特征就更为明显，需要我们在伦理审查等环节注重从程序法等角度去展开细致的制度设计。〔70〕

四、结　语

在当下，实务界各方亟待人工智能伦理合规制度能符合可操作、可预期、可计算的要求。通过对各种人工智能伦理原则内容的分析，可以发现几乎所有的人工智能伦理原则要求，都已经被世界主要国家和地区的数据、算法监管制度所涵盖。因此，时下惯常使用“人工智能伦理法治化”或者“人工智能伦理规范化”等泛化的概念，难以清晰而有效地作为人工智能伦理合规实操化的制度构建目标和分析工具。更甚之，这些概念还无法呈现或者遮蔽了人工智能立法中“安全与创新”“规范和发展”等诸多对立性讨论的问题意识。

法化概念的最初使用，应与当下中国语境中的“法治化”同义。但随着时间推移，“法治化”已经逐渐空泛化为类似于“规范化”“制度化”的概念。相比而言，通过法社会学理论脉络的不断充实，“法化”既具有丰富内涵的事实描述功能，而且还具有对权利证立、程序保障等“法治化目标”的反思功能。更重要的是，在前沿科技深远性影响社会实践的当下，通过法化的概念，还可以更好地分析一些传统并不属于传统正式法律制度的科技伦理、技术标准、企事业单位内部规范的法律性质。

科技伦理的既有法化路径主要有三：道义论持有者常用的新兴权利证立（比如人格权）、结果论持有者常用的软法化（比如政府监管工具创新）、美德论持有者常用的共同体伦理化（比如职业伦理、商业伦理）。那么，在上述研究的基础上，该如何在实操的角度，实现人工智能伦理的法化之路呢？显然，最便利的方式是通过与已经被高度法化的生物医药伦理相比较（当然，还有一种在生物医药伦理传统上进行续接的思路），抽象出人工智能伦理的特征，进而从中推导出适合人工智能特有的（规范性的）伦理要求。

法学方法论中的事物本质理论，主要的思路是在法律适用环节，“用事物本质的概念嫁接事实与价值的鸿沟”，近来虽在部门法理论研究也有一些应用性成果，由于其科学性程度不高的问题，整体上呈现的较为沉寂。〔71〕但在人工智能伦理法化的场景里，因广义意义上立法的民主化过程（商谈、利益博弈）的性质所决定，它可能是通过对人工智能技术特征及其引发的社

〔70〕 See Zhou and Chen, supra note 60, pp. 2693-2703.

〔71〕 近五年中文学术界部门法应用的研究文献可以参见熊伟：“事物本质、领域区分与领域法的特性透视”，《政法论丛》2024年第1期，第15—25页。

会关系变革，逐步推导出独特的伦理要求的有效理论工具。在此思路引导下，我们概括了人工智能伦理法化的三个约束性要件：道德规则的技术可嵌入性、更强的场景性以及依赖于技术过程的程序性。

前文大胆地将产生于多元价值观时代、具有鲜明批判传统的“法化”概念，与产生于一元化价值观时代、以面向法律适用为主要工具目的的“事物本质”概念融于一起，将其运用于人工智能伦理这一新兴事物的讨论中，在理论概念的深化、论证的严密性必然还有很多补充性的工作。更重要的是，当下的诸多人工智能研究还启示我们，关于人工智能和社会结合的研究，必须要回到复杂性的研究思路，才能获得更好的解决答案。〔72〕 比如，人工智能伦理应主要依赖于至下而上生成的思路，如果借用复杂适用系统可能会获得更好的分析。〔73〕 但限于篇幅对此无法再做进一步展开。

Abstract: Transforming AI ethics from moral principles into operable, predictable, and calculable ethical compliance practices, including ethical review requires exploring the legal nature of AI ethics, especially how the legal system evaluates and incorporates AI ethics rules. The two traditional socio-legal theories, “juridification” and “nature of thing”, can be used to as analytical tools. On the one hand, compared with overgeneralized concepts such as “legalization” and “standardization,” juridification offers better convergence when analyzing the practical implementation of AI ethics. The existing juridification paths for ethics of science and technology mainly involve three approaches: the justification of emerging rights (especially personal rights) common among deontologists, the soft law and innovative government regulation favored by consequentialists, and the professional ethics approach often employed by virtue ethicists. On the other hand, in deriving normative ethical requirements from the characteristics of AI technology and its industry, legislators or “legal discoverers” can incorporate the concept of nature of thing, abstracting three constraining elements of AI ethics legalization through comparison with biomedical ethics: the technical embeddability of moral rules, increased contextuality, and procedural dependence on technological processes.

Key Words: AI ethics; Ethics of Science and Technology; Ethical Review; Juridification; Nature of Thing

（责任编辑：王锡锌）

〔72〕 See Bickley and Torgler, supra note 22, pp. 501-519.

〔73〕 See Jakob Stenseke, “On the Computational Complexity of Ethics: Moral Tractability for Minds and Machines,” *Artificial Intelligence Review*, Vol. 57, No. 4, 2024, pp. 1-90; J. C. Gellers, “AI Ethics Discourse: A Call to Embrace Complexity, Interdisciplinarity, and Epistemic Humility,” *AI & Society*, Online First, 2023, pp. 1-2.