

人工智能立法的产品安全思路

——对欧盟《人工智能法》的批判性解读

何泽昊*

摘要 人工智能立法没有必然的路径。欧盟《人工智能法》的制定采取了“产品安全思路”，试图借助对产品安全立法的模仿、准用与调整来回应人工智能风险。产品安全思路的意义在于通过标准化进程来整合市场、通过锁定“最小成本规避者”来实现安全威慑，并通过公法规范的私法效力来激活产品责任体制。然而，人工智能的非标准化特征预示了标准化进程的困难，人工智能的多元使用场景揭示了义务分配的复杂性，信息成本高昂的事前标准也难以激活产品责任的运作。考虑到中欧之间迥异的市场结构、制度语境与产品安全思路自身的局限，我国并无采纳产品安全思路的必要。相反，我国应认真对待人工智能领域的不确定性，借助责任体制的信息生产功能为下一步法律动作提供事实依据。

关键词 人工智能法 欧盟 产品安全 技术标准

“人工智能是一种新电力”^{〔1〕}，人工智能的电力隐喻既预示着此类技术广泛的应用场景，也对人工智能立法的恰当定位提出疑问。正如人类社会并不存在一部将所有通电的事物纳入控制范围的“电力法”，一部涉及所有法律部门且无所不包的“人工智能法”也无存在的可能与必要。基于这一常识，人工智能立法不具有必然的路径，需要立法者审慎地框定问题，并选择恰当的工具。

欧盟《人工智能法》是全球第一部综合性的人工智能立法。对于其他法域而言，当务之急并非争夺立法焦点，而是冷静审视《人工智能法》的“人工”属性。对此，“根据人工智能系统可能产生的

* 中国人民大学法学院博士研究生。本文系国家社会科学基金重大项目“当代中国数字法学基本范畴体系研究”（项目批准号：23&·ZD154）的阶段性成果。

〔1〕 Keith McAleer, “AI is the New Electricity”: Insights from Dr. Andrew Ng, Berkeley Sutardja Center for Entrepreneurship & Technology (July 29, 2024), <https://sctet.berkeley.edu/ai-is-the-new-electricity-insights-from-dr-andrew-ng/>.

风险的强度和范围,调整此类规则的类型和内容”〔2〕的“风险进路”不足以解释该法在结构和内容上的诸多选择。“风险”被《人工智能法》界定为“发生危害的概率和危害的严重程度的组合”,〔3〕该法针对人工智能系统的风险分级却并非建立在严格的量化评估之上,所谓的风险进路更多被作为一种认知工具,用以书写基本权利保护与人工智能发展相统一的“欧盟故事”。〔4〕

事实上,《人工智能法》的制定离不开对欧盟产品安全立法的模仿、准用与调整。〔5〕本文尝试以《人工智能法》着墨最多的高风险人工智能系统为样本,对这一“产品安全思路”予以全面地拆解和评估。只有了解欧盟立法的初衷与选择,我国的人工智能立法才有可能避免将前者的“人工”属性当成是不言自明的道理,发展适合我国的立法策略。

一、产品安全思路拆解

(一) 欧盟产品安全立法鸟瞰

欧盟的产品安全体制可以被划分为“旧进路”“新进路”与“新立法框架”三个阶段。在旧进路阶段,欧共同体试图对特定产品需要遵守的技术规格予以详尽地规定,以化解成员国之间的贸易壁垒,此种笨拙的立法方式难以回应现实产品的复杂性而收效甚微。〔6〕第二阶段为新进路阶段,其始于欧共同体理事会于 1985 年通过的决议。〔7〕根据该决议,规定产品在欧共同体流通应满足的要求的协调性立法应仅限于规定产品的“基本要求”,此类基本要求一般采用宽泛的绩效或者功能描述,以避免旧进路立法所面临的技术困难。将协调性立法中的基本要求转化为具体技术规格的任务则被交给欧洲的标准化机构,此种规格被称为“协调性标准”。通过遵守协调性标准,协调性立法所覆盖的产品可以获得推定的合法性。新立法框架则始于欧洲议会、欧盟理事会的 768/2008/EC 号决定与 2008 年版本的《市场监管条例》(Market Surveillance Regulation),这也标志着欧盟的产品安全体制进入第三阶段。新立法框架继承了“基本要求+协调性标准”的立法模式,同时对市场监管机制与产品的“合格评定”进行了优化。

如今,除纲领性的 768/2008/EC 号决定外,欧盟的产品安全体制包含以下两组核心法律文件:第一组是针对特定产品,例如机械、电梯、玩具的协调性立法,以及为此类协调性立法未能覆盖的产品和要求提供安全网作用的《一般产品安全条例》(General Product Safety Regulation);第二组是辅助于上述协调性立法,针对产品监管、“认可制度”与“CE 标识”的立法。其中,2008 年版本《市场监管条例》中有关认可制度与 CE 标识的规定继续生效,有关市场监管的规定

〔2〕 See Artificial Intelligence Act, Rec. 26.

〔3〕 See Artificial Intelligence Act, Art. 3(2).

〔4〕 See Regine Paul, *European Artificial Intelligence “Trusted throughout the World”: Risk-based Regulation and the Fashioning of a Competitive Common AI Market*, 18 Regulation & Governance 1065, 1074–1075 (2024).

〔5〕 较早提出这一观点的文献,参见 Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*, 22 Computer Law Review International 97, 102–106 (2021).

〔6〕 See Harm Schepel, *The Constitution of Private Governance — Product Standards in the Regulation of Integrating Markets*, Hart, 2005, p. 63.

〔7〕 Council Resolution of 7 May 1985 on a new approach to technical harmonization and standards.

则被 2019 年更新的《市场监管条例》所替代。与《一般产品安全条例》的适用范围类似,如果其他协调性立法中没有目标相同的、以更具体的方式规范市场监管和执法的条款,2019 年《市场监管条例》就得以适用。

忽略不同协调性立法的差异,此类立法所覆盖的产品合法进入市场的典型路径可做如下简化:制造商应识别其产品适用的协调性立法与适用的基本要求,并通过协调性标准或其他方式遵守基本要求。此后,制造商应起草包含产品设计、制造、操作等方面信息的“技术性文件”,并根据特定协调性立法所规定的程序开展合格评定。如果相关产品通过了合格评定,那么制造商就可以做出“合格声明”,以承诺其产品的合法性,并在产品上加贴 CE 标识。满足上述要求的产品就可以被合法地投放市场或投入服务。

(二) 适用范围

高风险人工智能系统是《人工智能法》的重心,该法第 6 条将高风险人工智能系统划分为产品构成类的高风险人工智能系统与独立的高风险人工智能系统。前者的认定直接依赖于欧盟产品安全立法的现成分类。当人工智能系统构成协调性立法覆盖的产品或者产品部件时,该系统就属于高风险人工智能系统。

独立的高风险人工智能系统则是用于执法、移民、庇护、边境管制管理等领域的人工智能系统。这一类系统的识别虽不直接依赖于产品安全领域的现成立法,但与产品构成类的高风险人工智能系统相同,其需要满足的基本要求、其提供者需要履行的义务均以产品安全立法为模板。换言之,《人工智能法》事实上以规范产品的方式规范高风险人工智能系统,这一点在通用人工智能模型相关规范中同样有所体现。

(三) 要求和义务

《人工智能法》事实上采取了“基本要求+协调性标准”的模式。以高风险人工智能系统为典型,该法规定的基本要求包括风险管理系统、数据治理、技术性文件、记录保存等。^[8]在此基础上,《人工智能法》与其他协调性立法都采取了以《欧洲标准化条例》(Regulation 1025/2012 on European Standardization)为基础的协调性标准制定程序。根据这一程序,欧盟委员会早已在 2023 年 5 月便正式向欧洲标准化委员会(CEN)和欧洲电工标准化委员会(CENELEC)发出“标准化要求”,涵盖了前述各项基本要求。

《人工智能法》将义务主体统称为“操作方”,包含提供者、产品制造商、部署者、授权代表人、进口商与分销商。^[9]虽然《人工智能法》对提供者和制造商予以区分,但制造商的概念主要用于在特定情形下将制造商识别为提供者,或者方便《人工智能法》与欧盟产品领域协调性立法的衔接。^[10]作为参照,新立法框架下的《医疗设备条例》采取了制造商、授权代表人、进口商、分销商的义务主体分类,并采用“经济操作方”的概念统摄上述主体。^[11]分属两部法律的“提供者”与“制造商”概念并无本质区别,都是开发(制造或整体翻新)系统(设备),或在这之后将其投放市场或投入服务的法律实体。^[12]授权代表人、进口商、分销商的划分则源于经典的产

[8] See Artificial Intelligence Act, Art. 7-15.

[9] See Artificial Intelligence Act, Art. 3(8).

[10] See Artificial Intelligence Act, Art. 25(3), Art. 43(3).

[11] See Regulation (EU) 2017/745, Art. 2(35).

[12] See Artificial Intelligence Act, Art. 3(3); Regulation (EU) 2017/745, Art. 2(30).

品流通链条。不难发现,只有部署者,亦即人工智能系统的专业使用者属于《人工智能法》的特色。

《人工智能法》建立了以提供者为核心的义务体系,此处同样以《医疗设备条例》为参照。高风险人工智能系统的提供者与医疗设备的制造商都应首先确保基本要求的满足,并建立与维持风险管理系统、质量管理体系、制定技术性文件、进行合格评定、做出合格声明、加贴 CE 标识、与监管机构合作、对不合规的产品采取纠正措施、建立与维持上市后监控系统等。^[13] 相比之下,其他义务主体的主要功能在于核查与确保提供者(制造商)的义务履行,与市场监管机构开展合作,并对不符合要求的人工智能系统(医疗设备)采取纠正措施。^[14]

(四) 合格评定

1. 何为合格评定

合格评定是制造商证明产品是否符合特定要求的过程。要理解新立法框架下合格评定的制度构成,就必须引入其“模块化”的立法策略。768/2008/EC 号决定将合格评定的程序划分为“模块 A”到“模块 H”共 8 个模块,其中一些模块还存在变体(如 A1、A2)。不同协调性立法对模块的选择与组合以不同产品的性质为依据,例如产品复杂性的高低、生产规模的大小、制造商自身技术水平的强弱、设计环节与制造环节的风险差异等,并尽量降低相关主体的负担。^[15]

除由制造商或其独立的内设机构开展评估的程序模块,其他程序模块均需要第三方“合格评定机构”的参与,此类机构需要经过“通知”程序才能实际开展评定。所谓“通知”程序,是指某一成员国通知欧盟委员会和其他成员国,根据某项欧盟协调性立法,某一合格评定机构已被指定进行合格评定,且该机构满足前述欧盟协调性立法对此类机构的资源与能力要求。^[16] 据此,各成员国应当建立作为公共机构的“通知机关”,负责对合格评定机构的指定、通知与监管;^[17] 被通知的合格评定机构则获得“被通知机构”的身份与实际开展服务的资格。

为确保分散在成员国的通知机关以统一的方式考察合格评定机构,进行通知,并确保合格评定在整个欧盟范围内的公信力,欧盟还建立了认可制度。根据 2008 年《市场监管条例》第 5 条,成员国的“国家认可机构”负责证明合格评定机构具备开展特定合格评定活动的技术能力。认可的法律效果在于为合格评定机构所应具有的能力提供权威证明,不论是通知机关,还是接收通知的欧盟委员会和成员国,都将优先根据认可证书来评价合格评定机构是否具有作为被通知机构的能力。

2. 高风险人工智能系统的合格评定

《人工智能法》第 16 条规定,所有的高风险人工智能系统都应当通过合格评定才能被投放市场或投入服务,这一合格评定制度主要通过对产品安全立法的模仿、准用和调整加以实现。

首先是模仿。《人工智能法》中的合格评定制度遵循了新立法框架所设定的模板,在通知机关与被通知机构的资格、通知的程序、认可的效力等诸多方面均未偏离前文的介绍,仅仅在细节

[13] See Artificial Intelligence Act, Art. 16 - 22; Regulation (EU) 2017/745, Art. 10.

[14] See Artificial Intelligence Act, Art. 22 - 24; Regulation (EU) 2017/745, Art. 11 - 14.

[15] See European Commission, Commission notice — The “Blue Guide” on the implementation of EU product rules 2022 (2022/C 247/01), 80.

[16] Ibid., at 88.

[17] See Decision No 768/2008/EC, Annex I, Art. R14.

层面进行了调整。例如,被通知机构的人员通常应满足法定的能力要求,而《人工智能法》的类似规定则要求被通知机构的人员“具备与相关类型的人工智能系统、数据和数据计算有关的经验和知识”^[18]。

其次是准用和调整。《人工智能法》第 43 条并没有直接对 768/2008/EC 号决定所给出的程序模块进行组合,而是根据高风险人工智能系统的类型分别设置评定程序:一是作为产品或其部件的高风险人工智能系统。考虑到此类产品所对应的协调性立法附属于新立法框架,并已经根据产品特性对合格评定的程序模块进行了选择和组合,对高风险人工智能系统的合格评定也应准用相同的程序模块,不过应将高风险人工智能系统应满足的要求加入合格评定的内容。二是独立的高风险人工智能系统。《人工智能法》并未针对其使用新立法框架现成的程序模块,而是通过附件六和附件七分别设计了一个基于内部控制的评定程序与一个需要第三方机构参与的评定程序。其中,基于内部控制的评定程序适用于绝大部分独立的高风险人工智能系统,需要第三方机构参与的评定程序则适用于生物识别系统。当存在可适用的协调性标准或“通用规格”时,生物识别系统的提供者既可以选择基于内部控制的合格评定程序,也可以选择有第三方机构参与的合格评定程序;当相关协调性标准和通用规格不存在、未被提供者采用,或其本身存在限制时,提供者便只能适用第三方参与的合格评定程序。

(五) 执法体制

《人工智能法》对产品安全立法的依赖还体现在执法体制的设计当中。该法第 74 条规定,2019 年版本的《市场监管条例》适用于人工智能系统,2019 年条例中的“经济操作方”包含《人工智能法》中的“操作方”,2019 年条例中的“产品”则包含所有的人工智能系统。这意味着,除非《人工智能法》存在与 2019 年《市场监管条例》目的相同且更加具体的条款,后者所规定的执法权力与执法程序将得到适用。《人工智能法》确实为非产品构成类的人工智能系统设置了专门的执法程序,但该程序仍旧以 2019 年《市场监管条例》所确立的执法程序为原型。^[19] 这一准用带来的问题是:如何为不同类型的人工智能系统确立市场监管机构? 在欧盟产品安全领域的执法权力由各成员国的市场监管机构所掌握的背景之下,《人工智能法》最大程度上尊重了这一现存的权力分配,并在特定情形中将数据保护领域的监管机构、欧盟委员会内设的“人工智能办公室”拟制为市场监管机构。

二、产品安全思路的意义与局限

(一) 意义

《人工智能法》为什么要以产品安全立法为原型? 起草者认为,产品安全领域的新立法框架能够为迥异的人工智能系统提供统一的法律框架,并且将软件作为产品或产品部件的立法尝试已经在欧盟产品安全领域践行了 15 年之久。^[20] 然而,此种辩护并不涉及产品安全思路自身的制度细

[18] See Decision No 768/2008/EC, Annex I, Art. R17(7); Artificial Intelligence Act, Art. 31(11).

[19] See Artificial Intelligence Act, Art. 74(4), Art. 79.

[20] See Gabriele Mazzini & Salvatore Scalzo, *The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts*, SSRN (July 29, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4098809.

节与政策考量。本文认为,人工智能立法的产品安全思路至少存在三方面的理据:通过协调性标准实现市场整合、通过锁定“最小成本规避者”实现安全威慑以及通过责任体制的激活实现损害救济。

1. 市场整合:利用协调性标准

要实现市场整合的目标,《人工智能法》的制定必须克服三方面的阻碍。首先,人工智能系统正在经历历史无前例的技术演进,通过立法直接对其应满足的技术规格加以固定既超出了立法机构的能力范围,也将挫伤人工智能开发者安全投入的积极性;其次,全球各大标准化机构早已进入人工智能相关技术标准的“军备竞赛”当中,人工智能领域的市场整合以技术标准的整合为前提;最后,即便是在欧盟内部,不同成员国对人工智能立法的态度也存在着分歧。例如,法国、德国和瑞典在过去十年中吸引了人工智能领域的大量投资,致使这些国家在人工智能立法的偏好上与其他国家产生差异。^[21] 法国和德国就曾担忧该法草案有关通用人工智能模型的规定将伤害其在相关领域的领先企业。^[22]

以此观之,“基本要求+协调性标准”立法模式无疑对《人工智能法》的起草者具有特别的吸引力。以基本要求为内容的立法避免了技术问题对立法进程的拖累,并能够在最大程度上凝聚成员国之间的政治共识。与此同时,协调性标准的制定不仅能够实现技术标准的统一,还将棘手的政治分歧转化为看似中立的科学问题。^[23]

2. 安全威慑:以提供者为核心

如前章所述,《人工智能法》以产品安全立法为师,将与制造商功能对等的提供者作为主要的义务主体,并通过分销链条下游主体的义务设置来确保提供者的可追溯性与可问责性。如果说产品安全立法的义务体系事实上将制造商视为“最小成本规避者”,处在对“事故成本及其预防成本进行识别并据此行动”的最佳位置上,那么《人工智能法》便共享了此种思维方式,将提供者视为人工智能风险的关键控制点。^[24] 在以自动驾驶为代表的部分人工智能应用场景中,这一最小成本规避者理论颇具解释力。随着人类驾驶员对车辆的控制力被转移到自动驾驶系统手中,开发、测试、安装自动驾驶系统,并能够根据真实世界的数据对自动驾驶系统进行安全更新的提供者无疑处在降低风险的恰当位置。^[25]

欧盟的产品安全立法之所以要对分销链条下游的授权代表人、进口商、分销商施加上游核查、执法合作等义务,正是考虑到制造商身处海外、执法机构鞭长莫及的情况。对人工智能产业相对弱势、进口人工智能系统不可避免的欧盟而言,《人工智能法》显然具有类似的忧虑。借助对欧盟内部节

[21] See Ronit Justo Hanani, *The Politics of Artificial Intelligence Regulation and Governance Reform in the European Union*, 55 Policy Sciences 137, 152 - 153 (2022).

[22] See Gian Volpicelli, *EU Countries Strike Deal on Landmark AI Rulebook: The Bloc's Law to Regulate AI Overcomes Threats of Late Opposition*, POLITICO (July 29, 2024), <https://www.politico.eu/article/eu-countries-strike-deal-ai-law-act-technology/>.

[23] See Harm Schepel, *The New Approach to the New Approach: The Juridification of Harmonized Standards in EU Law*, 20 Maastricht Journal of European and Comparative Law 521, 522 - 523 (2014).

[24] See Guido Calabresi & Jon T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 The Yale Law Journal 1055, 1960 (1972).

[25] See Mark A. Geistfeld, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, 105 California Law Review 1611, 1632 - 1660 (2017).

点的影响,执法机构能够影响到海外提供者的行为,进一步巩固了以提供者为中心的义务体系。

3. 损害救济: 激活产品责任体制

除了产品安全立法,产品责任立法同样是回应产品风险的重要制度工具。当产品缺陷导致消费者或第三方损害时,产品责任立法通常对产品的卖方尤其是制造商施加无过错责任。^[26]但在人工智能致害的场景中,产品缺陷的认定存在着不确定性。以设计缺陷为例,不论是将设计缺陷理解为对消费者正当期望的背离,^[27]抑或未能采取具有风险效用合理性的替代设计,^[28]均存在着困难。一般消费者的期望难以在短时间内形成,人工智能技术的复杂性亦令法院难以完成风险效用评估的任务。

借助与《人工智能法》的配合,欧盟于2024年通过的新《产品责任指令》在一定程度上回应了人工智能致害场景中的产品缺陷认定问题。新《产品责任指令》第10条特别规定,如果原告表明被告未能履行欧盟立法或成员国立法中旨在回应原告所受损害的产品安全标准,那么就可以推定产品缺陷的存在。这意味着,当人工智能系统构成产品或其部件时,《人工智能法》及相关协调性标准就能够降低产品缺陷认定的难度。

(二) 局限

1. 人工智能的标准化难题

技术标准最早以19世纪的蒸汽锅炉、螺纹、钢轨等工业部件为对象,并协助了当时的工业大规模生产。^[29]不可否认,技术标准并非仅仅适用于工业化生产的产品与产品部件,而是具有极强的泛用性。但就人工智能立法所欲实现的目的与人工智能系统的技术特性而言,针对人工智能系统的标准化困难重重。

除了追求安全、健康等产品领域的惯常目标,《人工智能法》将基本权利保护作为自身的根基。然而,与健康风险、安全风险这些可以通过死亡率、致病率、预期寿命减少等指标加以量化的产品风险相比,人格尊严等方面的基本权利风险具有丰富的价值诠释空间,难以被简单度量与验证,相应的标准化过程将不可避免地存在模糊、武断与错置。^[30]延伸而言,基本权利的标准化困境至少源于以下两方面事实:其一,价值领域不存在所谓的专家。以欧盟为例,该法域近年来围绕ICT治理出台了一系列伦理评估的框架,授权伦理委员会、咨询小组等机构开展专家主导的评估程序。但论者指出,相关评估框架缺乏坚实的科学基础,所谓“伦理专家”的知识背景也五花八门,同行评议缺位;^[31]其二,对价值问题的解决呼唤着更多的民主输入,而技术标准的民主正当性却在承受愈来愈多的质疑。特别是,欧盟的协调性标准并非法律,却比产品安全立法更有力地影响着市场

^[26] 此类产品责任立法的全球传播,参见 Mathias Reimann, *Product Liability in a Global Context: The Hollow Victory of the European Model*, 11 *European Review of Private Law* 128, 130-133 (2003).

^[27] See Daily Wuyts, *The Product Liability Directive — More than Two Decades of Defective Products in Europe*, 5 *Journal of European Tort Law* 1, 8-13 (2014).

^[28] See Aaron D. Twerski & James A. Henderson, Jr., *Manufacturers' Liability for Defective Product Designs: The Triumph of Risk-Utility*, 74 *Brooklyn Law Review* 1061, 1062-1071 (2009).

^[29] See JoAnne Yates & Craig N. Murphy, *Engineering Rules: Global Standard Setting since 1880*, Johns Hopkins University Press, 2019, p. 28-34.

^[30] See Marco Almada & Anca Radu, *The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy*, 25 *German Law Journal* 1, 7-8 (2024).

^[31] See Niels van Dijk, Simone Casiraghi & Serge Gutwirth, *The “Ethification” of ICT Governance. Artificial Intelligence and Data Protection in the European Union*, 43 *Computer Law & Security Review* 1, 11-12 (2021).

主体的行为。为此,《欧洲标准化条例》加强了中小企业、消费者组织等社会利益相关方的正当性输入,但这一程序仍无力帮助协调性标准获取如欧盟立法一般的广泛代表性。^[32]与确立 A4 纸张大小等纯粹技术事项不同,基本权利方面的决策以利益代表的恰当构成与偏好凝聚为前提,^[33]传统上由专家主导的标准制定程序因此存在先天不足。

更为关键的是,标准化可能并不符合人工智能的技术特点。对于欧盟之外的法域,暂时搁置基本权利问题并专注于生命、健康、财产等有型风险亦是一种可考虑的立法策略。在此种情形下,人工智能系统的标准化是否就有可能实现?答案是否定的。当下的人工智能技术以机器学习模型,特别是人工神经网络模型为基础。在人工神经网络模型的开发过程中,开发者对神经网络结构、超参数、学习算法等方面的选择常常依赖于反复尝试及由此形成的经验法则,而非精确的科学。^[34]无怪乎,大量开发者将人工神经网络的超参数调整称为“炼丹”,只能在成功与失败的循环中拿捏恰当的火候。要通过标准化为人工智能系统提供可预期的安全性,流动的经验就需要转化为精确的科学知识,但在人工智能的“炼丹”时刻,这显然是一个艰巨的任务。^[35]

2. 义务分配的场景复杂性

《人工智能法》的义务框架更像是为产品构成类的人工智能系统量身打造,对于其他人工智能系统而言则颇有生搬硬套之感。例如,对于公权力机关委托研发或自主研发的独立人工智能系统,所谓的进口商、授权代表人与分销商并不存在。即便仅仅将目光局限在产品构成类的人工智能系统,以提供者为中心的义务框架仍存在着理论难以自治之处。概言之,提供者常常并非唯一能够以合理成本预防人工智能风险的主体。究竟哪些主体能够以合理成本预防人工智能风险?对这一问题的回答至少需要纳入以下两方面的考量:

首先,谁更有可能创造风险。高度自动驾驶系统的提供者之所以构成典型的最小成本规避者,其原理恰好在于将人类驾驶员创造风险的能力降至最低,既然风险更有可能来自自动驾驶系统本身,要求其提供者采取安全措施便具有充分的理由。^[36]然而,另一类人工智能系统则增强了系统使用者创设风险的可能,其中最为典型的例子是所谓“深度伪造”技术的应用。其实,通过数字技术捏造事实并非新现象,但深度伪造技术及基于此种技术传播的应用程序则大大降低了普通人作恶的门槛。^[37]对于此类人工智能系统,风险的产生同时来自提供者的系统设计与使用者的使用方式,提供者的单方安全措施难以回应所有风险,对提供者施加过高的义务甚至可能对人工智能系统的善意使用造成妨碍。

[32] See Rob van Gestel & Hans-W Micklitz, *European Integration through Standardization: How Judicial Review is Breaking down the Club House of Private Standardization Bodies*, 50 *Common Market Law Review* 145, 177 - 181 (2013).

[33] 参见赵鹏、谢尧雯:《科技治理的伦理之维及其法治化路径》,载《学术月刊》2022 年第 8 期。

[34] See Bryan H. Choi, *AI Malpractice*, 73 *DePaul Law Review* 301, 310 - 330 (2024).

[35] 越是具有不确定性的领域,越难以制定合理的事前标准。参见宋亚辉:《环境管制标准在侵权法上的效力解释》,载《法学研究》2013 年第 3 期,第 47—48 页。

[36] See Gerhard Wagner, *Robot Liability*, in Sebastian Lohsse, Reiner Schulze, Dirk Staudenmayer eds., *Liability for Artificial Intelligence and the Internet of Things: Münster Colloquia on EU Law and the Digital Economy IV*, Hart, 2018, p. 37 - 39.

[37] See Bobby Chesney, Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 *California Law Review* 1753, 1762 - 1768 (2019).

其次,谁更有可能了解风险。以提供者为核心的义务框架假定,实际开发人工智能系统的提供者对于人工智能系统产生风险的方式具有最全面的认识。然而,人工智能系统所产生的风险不仅仅源于技术本身,还源于其应用场景。例如,医学领域的管理与诊疗需求在很大程度上取决于特定地区与医院所面向的患者群体、已经建立的工作流程、医生的训练水平等复杂因素,医疗人工智能系统的提供者即便以服务不特定的患者为目标,其训练数据集仍然不太可能覆盖所有地区与所有医院的特殊情况。此时,对地方医疗条件具有全面认知的医院就能够发挥医疗人工智能系统的守门人角色,在此类系统的采购与测试过程中降低“水土不服”所可能造成的风险。^[38]此外,当人工智能系统属于产品部件时,相较于作为“部件制造商”的提供者,终端产品的制造商显然更加了解终端产品的应用场景。^[39]同样是具有会话功能的人工智能系统,当被安装在成年人使用的智慧家电和儿童专属的玩具上时,所应采取的安全措施便完全不同。

值得强调的是,《人工智能法》对上述问题并非全然忽视,但其回应方式却是不完整的。例如,当人工智能系统被用于深度伪造,作为专业使用者的部署者应当披露其生成的内容属于人为生成或操纵。^[40]常识告诉我们,非专业使用者才是民间恶意萌发的源头,将使用者责任局限在专业使用者并不合理;又如,对于人工智能系统构成产品部件的情形,该法将终端产品的制造商也认定为提供者,只要该制造商以自己的名称或商标将前述系统与终端产品共同投放市场或投入服务。^[41]这一规定有助于激励终端产品的制造商在人工智能系统部件的开发环节便提供场景化的知识,弥补系统开发者自身的知识盲区。但正如前文有关医疗人工智能系统的论述,具有宝贵地方性知识的主体往往是医院等中间角色。总之,尝试为人工智能治理提供统一框架的《人工智能法》无力在复杂场景中实现合理的义务分配。

3. 事前标准的信息劣势

在立法者的眼中,《人工智能法》与协调性标准一类的事前标准或将成为产品责任体制的助力,为人工智能时代的产品缺陷认定提供确定性,此种设想忽视了事前标准制定所需的高昂信息成本。前述标准化难题与义务分配的复杂性均是这一信息困境的侧写。

事前标准的信息劣势源于以下三方面的事实。其一,人工智能领域较高的创新频率。相较于药品研发等领域,人工智能领域的研发成果能够以更低成本被反复使用,下游研发人员还能够根据上游研发人员的工作成果进行“微调”式创新。^[42]其二,人工智能领域的信息不对称。相比公共财政支持的研究机构,私人领域吸收了绝大多数的人工智能专家,汇聚了绝大部分算力,掌握着规模最大的人工智能模型,并在大多数应用领域的基准测试中占据领先地位。^[43]缺乏必要信息

[38] See W. Nicholson Price II & I. Glenn Cohen, *Locating Liability for Medical AI*, 73 DePaul Law Review 339, 360 (2024).

[39] 参见丁晓东:《人工智能风险的法律规制——以欧盟〈人工智能法〉为例》,载《法律科学》2024年第5期,第16页。

[40] See Artificial Intelligence Act, Art. 50(4).

[41] See Artificial Intelligence Act, Art. 25(3).

[42] See Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 Southern California Law Review 633, 653 - 656 (2020).

[43] 相关数据,参见 Nur Ahmed, Muntasir Wahed & Neil C. Thompson, *The Growing Influence of Industry in AI Research: Industry is Gaining Control over the Technology's Future*, 379 Science 884, 884 - 886 (2023).

的公共机构难以制定合理的事前标准。其三,人工智能领域自身的信息匮乏。即便是在大量“热钱”涌入人工智能领域的当下,人类仍未清晰掌握前沿模型“举一反三”的原理。^[44] 针对前沿人工智能系统的事前标准缺乏可靠事实的支撑。据此,事前标准先行从而带动责任体制的做法并不可取。过于精细的事前标准可能对行业创新施以沉重的打击,过于模糊的事前标准则难以为责任体制提供确定性。

三、产品安全思路与中国道路

(一) 我国无须采取产品安全思路

对于我国的人工智能立法而言,产品安全思路具有技术上的可行性。尽管在法律的分割与编排上与欧盟差异极大,我国的产品安全立法却也包含下述制度模块:其一,基于技术标准的产品要求。根据《产品质量法》第13条,可能危及人体健康和人身、财产安全的工业产品必须符合相关国家标准、行业标准。其二,完整的认证认可制度。根据《认证认可条例》第2条,认证制度包含对产品的合格评定,而认可制度则包含对认证机构及相关从业人员的能力和执业资格的承认。特别是,《强制性产品认证管理规定》建立了强制性产品认证制度,根据“保护国家安全、防止欺诈行为、保护人体健康或者安全、保护动植物生命或者健康、保护环境”的需要,特定产品“必须经过认证,并标注认证标志后,方可出厂、销售、进口或者在其他经营活动中使用”。此类产品应适用的认证规则被划分为数个模块,市场监管总局根据产品的风险特性对这些模块予以选择和组合。其三,我国同样建立了以市场监管机构为担纲者的产品执法体制,授予其包含召回在内的广泛执法权力。

然而,产品安全思路对于我国人工智能立法的政策价值却非常有限。首先,人工智能系统难以被简单标准化,人工智能系统复杂的应用场景使得以提供者为中心的义务体系捉襟见肘,囿于信息困境的事前标准也无法为责任体制提供助力。对于任何采纳产品安全思路的法域而言,上述问题都无回避的可能性,这构成了我国反对产品安全思路最重要的理由。其次,我国并不具有市场整合的强烈需求。我国虽然存在地方保护主义等形式的经济壁垒,但是作为一个主权国家,我国内部仍存在一个相对统一的市场,对经济壁垒的消除并不依赖于产品乃至人工智能领域的统一立法。最后,我国能够选择的立法策略远比欧盟丰富。欧盟机构不具有独立的财政权力,因此由企业和成员国承担法律运行成本的规制性立法是欧盟机构追求其政策目标最有力的手段。^[45] 我国的人工智能立法则并不受到上述制度的限制,完全可以采取产业促进型的立法,或对侵权法、知识产权法等法律部门进行调整。

(二) 责任体制先行:一种务实的立法策略

我国人工智能立法应当采取何种策略? 当前研究在统一立法还是分散立法、着重何种法律部

[44] See Will Douglas Heaven, *Large Language Models Can Do Jaw-Dropping Things. But Nobody Knows Exactly Why*, MIT Technology Review (July 29, 2024), <https://www.technologyreview.com/2024/03/04/1089403/>.

[45] See Anu Bradford, *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020, p. 16-18.

门、安全与发展的价值如何平衡等诸多问题上仍然争论不休。^[46] 从 2023 年开始,国务院连续两年在年度立法工作计划当中提请全国人大常委会审议人工智能法草案,全国人大常委会于 2023 年 9 月发布的五年立法规划中却并未将人工智能立法纳入规划。可见,我国的人工智能立法进程同样显现出反复、审慎的姿态。人工智能立法所面临的不确定性来源于人类社会的信息赤字,前沿人工智能模型的技术原理为何、人工智能系统究竟会产生何种具体风险,此种风险又是否令现有法律到了非改不可的境地? 这些问题均难以在短期内获得答案。

面对信息赤字,责任体制可以在个案中生产一系列“高颗粒度”的信息,从而为规制标准、市场机制的建立与优化提供帮助。长期以来,学者将侵权法的功能局限在补偿、威慑、实现矫正正义、提供“民事救济”(civil recourse)等方面,^[47]却极少关注到侵权法的信息生产功能。申言之,特定的责任体制,如过失责任与基于风险效用测试的产品责任均能够在个案中生产一系列“高颗粒度”的信息,如损害的性质、数量与发生频率,原被告行为在损害发生中起到的作用,特定领域的技术水准与行业惯例等。^[48] 例如,当自动驾驶系统未能识别特定障碍物并导致乘客损害时,责任体制需要考虑:障碍物的识别失败是否源于设计缺陷、自动驾驶系统未能识别障碍物时是否向乘客发出接管车辆的提示、此种提示是否足以引起乘客注意,以及乘客是否无视了这一提示等等。这些基于真实损害的信息能够帮助消费者作出知情的消费决策,并为事前标准的制定提供广泛的技术与社会基础,避免基于想象乃至恐慌的标准制定。^[49]

当然,通过责任规则实现信息生产的方案需要回应诉讼成本的挑战。在人工智能的“黑箱”语境下,当事人的举证难度将会增加,法院也将难于确立责任成立的主客观要件。但如果仔细考量,就会发现相关诉讼成本相对可控,其存在也具有社会公益上的正当性。之所以说其相对可控,是因为人工智能致害场景中的过错、产品缺陷、因果关系等主客观要件常常能够根据间接证据加以推断,而无须探究复杂的技术细节。^[50] 例如,当自动驾驶汽车无理由地突然加速并造成人身损害时,突然加速的事实本身就足以推断产品缺陷的存在。此类推断之所以能够成立,是因为诸多人工智能应用是传统产品与服务的智能化改造,^[51]前人工智能时代所形成的经验法则就足以帮助当事人与法院判断责任是否成立。此外,过错推定与举证责任倒置能够将举证负担施加给人工智能系统的提供者或专业使用者,同样可以降低原告的诉讼成本。诉讼成本存在的正当性则源于人工智能相关信息作为公共益品的属性。“益品”源于此类信息降低市场交易成本,提高政府规制效能,从而增进社会利益的特质,“公共”则是说此类信息一经产生便具有非竞争性与非排他性。理

[46] 代表文献,参见张凌寒:《中国需要一部怎样的〈人工智能法〉?——中国人工智能立法的基本逻辑与制度架构》,载《法律科学》2024年第3期;周汉华:《论我国人工智能立法的定位》,载《现代法学》2024年第5期。

[47] 参见程啸:《侵权责任法(第三版)》,法律出版社2021年版,第30—40页;See Guido Calabresi, *The Costs of Accidents: A Legal and Economic Analysis*, Yale University Press, 1970, p. 24-33; John C. P. Goldberg & Benjamin C. Zipursky, *Recognizing Wrongs*, Harvard University Press, 2020, p. 25-178.

[48] See Assaf Jacob & Roy Shapira, *An Information-Production Theory of Liability Rules*, 89 *The University of Chicago Law Review* 1113, 1126-1136 (2022).

[49] 参见[美]史蒂芬·布雷耶:《打破恶性循环:政府如何有效规制风险》,宋华琳译,法律出版社2009年版,第42—70页。

[50] See Bryan Casey, *Robot Ipsa Loquitur*, 108 *The Georgetown Law Journal* 225, 270-273 (2019).

[51] See Jack B. Balkin, *The Path of Robotics Law*, 6 *California Law Review Circuit* 45, 55-59 (2015).

性主体,如畏惧于司法成本而放弃诉讼的潜在原告缺乏提供此类信息的动机。在此种意义上,高昂的诉讼成本正是生产公共品所必要的投入。上述辩护同时也预示了人工智能致害场景中责任体制的改革方向,即由国家来承担部分诉讼成本。例如,可以设立专门基金,对人工智能致害场景下满足特定条件的自然人原告进行诉讼补贴。从全球范围看,此种为专门目的设立的诉讼基金也并非一个全新的实践。^[52]

四、结 论

人工智能立法没有必然的路径。作为全球首部综合性的人工智能立法,《人工智能法》的“人工”属性应被其他法域的立法者所清晰认知。该法采取了“产品安全思路”,尝试通过对产品安全立法的模仿、准用与调整来回应人工智能时代的风险问题。然而,产品安全思路难以解决人工智能系统在标准化方面的困难,难以在复杂应用场景中合理地分配义务,也难以通过信息成本高昂的事前标准为责任体制提供帮助。我国应充分认识到人工智能领域的高度不确定性,通过责任体制的运作为下一步法律行动提供事实基础的方案应被认真考虑。

Abstract AI legislation does not have an inevitable path. The European Union's AI Act adopts a “product safety approach” in its legislative strategy, attempting to respond to the risks of AI through the imitation, application and adjustment of product safety legislation. The significance of the product safety approach is to integrate the market through the process of standardization, to achieve safety deterrence by locking the “Cheapest Cost Avoider”, and to activate the liability system through the public law. However, the non-standardized nature of AI indicates the difficulty of the standardization process, the diversity of AI use cases reveals the complexity of obligation allocation, and the information-costly ex ante standards do not provide an “antidote” to the liability regime. Given the significant differences in market structures and institutional contexts between China and the EU, as well as the limitations of the product safety approach itself, there is no need for China to adopt the product safety approach. Instead, we should take the uncertainty in the field of AI seriously and use the liability system's information production function to provide a factual basis for the next legal action.

Keywords AI Legislation, EU, Product Safety, Technical Standards

(责任编辑:曹博)

^[52] See Maya Steinitz, *Whose Claim Is This Anyway? Third-Party Litigation Funding*, 95 Minnesota Law Review 1269, 1275 - 1285 (2011).