欧盟《人工智能法》：演进、规则与启示*

刘子婧

摘　　要： 欧盟《人工智能法》遵循“政策—软法—硬法”的演进脉络，是欧洲数字单一市场战略的法律化产物，旨在实现整合数字市场、促进创新发展、防控社会风险和保障基本权利多重立法目标。由于立法目标间不可避免地存在内在张力，在“基于原则”“基于权利”和“基于风险”三种人工智能法律规制路径中，立法者最终选择了“基于风险”的规制路径，通过强化企业负责性而非赋予公民个体控制性权利的方式实现监管与创新的平衡。在以风险为中心的立法格局下，该法为人工智能系统设计了禁止、合规和豁免规则，以期实现监管的可伸缩性和合比例性。这部立法体现了一种“规制-工具主义”思路，在延续《通用数据保护条例》中“基于风险”方法的同时，也承继了《产品责任指令》中的产品监管框架。我国应当辩证地看待欧盟人工智能立法的成败得失，在吸收其立法经验的基础上，审慎地思考我国未来人工智能立法的方向。

关 键 词： 欧盟《人工智能法》；　人工智能立法；　风险；　权利

作者简介： 浙江大学　光华法学院　博士研究生　杭州　310008

中图分类号： TP18；　D99

文献标识码： A

文 章 编 号： 1005－4871(2024)03－0101－28

* 本文系浙江大学“重要国家和区域研究”专项“欧盟人工智能立法研究”(项目编号：S20240070)的阶段性成果。

引 言

由大数据、云计算和人工智能(Artificial Intelligence)掀起的技术革命浪潮被视为“第四次工业革命”,在这场技术革命中,没有什么比人工智能技术更引人瞩目。[①] 如今,人工智能已无处不在,创造出巨大的商业价值和市场潜力,成为世界经济增长新引擎,对整个人类社会都产生深远影响。据预测,到2030年,大约70%的公司将采取至少一种人工智能技术,人工智能将给全球国内生产总值(GDP)带来14%的增长。[②] 同时,人工智能也给以人类智能为中心的法秩序带来前所未有的“破窗性”挑战和“创造性破坏”,如何对其进行法律规制是全球立法者与学术界共同关注的议题。[③] 在这场有关人工智能法律规制的争论中,欧盟立法者仍保持其“全球数字规则制定者”的声誉。欧盟委员会于2021年发布了欧盟《人工智能法(草案)》(Commissions Proposal for Artificial Intelligence Act),2024年5月21日,在经过多轮讨论与修改后,欧盟理事会正式批准通过《人工智能法》(Artificial Intelligence Act),该法成为世界上第一部专门针对人工智能系统的横向集中式立法。[④] 欧盟《人工智能法》的出台并非易事,它既要追赶快速变化的技术发展,又面临着多重法益的复杂平衡。一方面,人工智能技术尚处于质变前夜,立法需要追赶

---

① Klaus Schwab, “The Fourth Industrial Revolution: What It Means and How to Respond”, World Economic Forum, 2016-01-14, https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-ho-to-respond/,访问日期:2024-03-20。

② Ethan Ilzetzki/Suryannsh Jain, “The Impact of Artificial Intelligence on Growth and Employment”, 2023-06-20, https://cepr.org/voxeu/columns/impact-artificial-intelligence-growth-and-employment,访问日期:2024-02-20。

③ 国内讨论参见郑志峰:《人工智能的法律挑战与规制重点》,载《月旦民商法杂志》,2022年第6期,第18-27页;汪庆华:《人工智能的法律规制路径:一个框架性讨论》,载《现代法学》,2019年第2期,第54-63页;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(西北政法大学学报)》,2024年第3期,第3-17页;赵精武:《论人工智能立法的多维规制体系》,载《法学论坛》,2024年第3期,第53-66页。域外讨论参见Dane Chapman, “The Ideal Approach to Artificial Intelligence Legislation: A Combination of the United States and European Union”, *University of Miami Law Review*, Vol. 78, No. 1, 2023, pp. 256-296; Thomas Wischmeyer/Timo Rademacher (eds.), *Regulating Artificial Intelligence*, Cham: Springer, 2020, pp. 1-32。

④ 本文所称的《人工智能法(草案)》指2021年4月通过欧盟委员会提案版本,《人工智能法》指2024年4月19通过的版本,该版本是欧洲议会对其3月13日批准通过的条例进行语法和数字方面的订正后的最终版本。参见European Commission, “Proposal for Artificial Intelligence Act”, COM (2021) 206 final, 2021-04-21,https://eur-lex.europa.eu/legal-content/EN/TXT?uri=celex%3A52021PC0206, 访问日期:2024-04-22。European Commission, “Artificial Intelligence Act — Corrigendum to the Position of the European Parliament Adopted at First Reading on 13 March 2024”, COM(2021)0206-C9—0146/2021-2021/0106(COD), 2024-04-19,https://artificialintelligenceact.eu/ai-act-explorer/,访问日期:2024-04-22,若无特别说明,本文所使用条例版本均为此版,后续引用时简称《人工智能法》。

上技术发展的步伐；另一方面，立法需要面对技术控制的“科林格里奇困境”(Collingridge Dilemma)[①]，小心平衡监管与创新间的紧张关系。

作为世界上第一部人工智能立法，欧盟《人工智能法》无疑具有开创性意义。其迎接了高新技术对传统立法格局的挑战，进一步发展了《通用数据条例》(General Data Protection Regulation，GDPR)中“基于风险”的规制路径，致力于通过这种可伸缩、合比例的监管方式实现创新和监管的二元平衡。此外，这部立法还将对世界人工智能监管格局产生重要影响：一则，基于先发优势，欧盟《人工智能法》将成为其他国家人工智能立法的参照系，对全球监管产生间接影响；二则，借助立法中的域外效力条款和欧盟庞大的市场体量，欧盟法规的“布鲁塞尔效应”(Brussels effect)[②]可能会再次出现，并对全球监管产生直接影响。总之，欧盟《人工智能法》凝聚了欧洲社会共识，代表了欧盟立法者在人工智能法律规制问题上的价值取向与立法选择，可能成为全球人工智能立法标杆。但是，作为一项实现“零突破”的立法尝试，欧盟采取的路径是否代表了人工智能法律规制最佳方案还有赖细致的分析和检讨。本文首先回顾欧盟人工智能立法的历史脉络，沿着政策与立法演进的轨迹，探析欧盟立法者的立法考量，并试图还原欧盟数字立法的动态图景。其次，本文对欧盟立法过程中呈现的“基于原则”(principle-based)、“基于权利”(rights-based)和“基于风险”(risk-based)三类不同立法路径进行抽象提炼和比较分析，考察欧盟立法者为何最终选择“基于风险”的法律规制思路，并梳理在以风险为中心的立法构造下，这部立法的基本框架与核心制度。最后，本文对欧盟人工智能立法进行反思与评析，讨论其对我国人工智能立法的启示与借鉴，并对我国未来人工智能立法提出具体建议。

一、欧盟《人工智能法》的演进历史

立法演进的轨迹揭示了欧盟《人工智能法》与“数字欧洲议程”(Digital Agenda for Europe)之间的内在关联，借助历史的线索，我们可以更好地理解该法诞生的内在动因。欧盟人工智能立法大致分为三个阶段：政策形成期(2010—2016 年)、软法时期(2017—2019 年)和硬法时期(2020—2024 年)。

① 该困境由技术哲学家大卫·科林格里奇在其 1980 年所著《技术的社会控制》一书中提出，是指技术控制中存在的一种二难推理，即“如果因担心危害后果而过早实施技术控制，则技术革命难以爆发；而如果过晚实施控制，则技术已经成为整个经济和社会结构的一部分，并且走向失控的边缘，此时的控制成本过于高昂，困难且耗时”。Collingridge David，*The Social Control of Technology*，London：Frances Pinter，1980，p. 11.

② “布鲁塞尔效应”由阿努·布拉德福德(Anu H. Bradford)提出，指欧盟凭借其市场力量单边监管全球市场的一种效果。Anu Bradford，*The Brussels Effect*：*How the European Union Rules the World*，New York：Oxford University Press，2020，pp. 1－6.

(一)“欧洲数字议程”与立法版图的初构

欧盟《人工智能法》诞生于欧洲“数字单一市场”(Single Digital Market)战略的关键时期,是欧盟“欧洲数字议程”的法律化产物。[①] 早在2010年,随着信息与通信技术的勃兴,为最大化地激发信息技术所带来的社会和经济效益,欧盟委员会开始全面制定“欧洲数字议程”,包括:第一,建设“数字单一市场”,解决欧盟数字市场碎片化问题,促进商品、服务和文化产品的在线流动;第二,促进信息、通信等数字技术的投资与创新,为成员国带来经济增长;第三,提升公共部门与私人的数字素养和技能,包括数字政府建设等。[②] 其中,建设“数字单一市场”成为“欧洲数字议程”的重中之重,也是数字时代下欧洲一体化战略的重要组成部分。2015年,欧盟委员会发布“数字单一市场”战略,旨在建立一个任何企业与个人均可进行无线访问和在线活动的“数字单一市场”。[③]

“数字单一市场”建设有赖于监管协同,后者以统一的数字经济规则和立法为基础。数字资源(例如数字产品、技术和服务)与传统资源(例如能源、土地)的一项重要区别是:前者更多的是由超大型跨国公司掌握和提供,而跨国企业具有全球化和脱嵌性特征,需要超国家的政治机构来为数字经济制定规则。因此,在雄心勃勃的“数字单一市场”战略和数字监管协同的目标下,欧洲立法者开始部署数字领域的统一立法,以消除跨境在线贸易、数字基础设施服务、知识产权与个人信息保护方面的法律壁垒,建立中立、兼容、标准的法律和技术规范,促进要素自由流通和市场公平竞争。[④]

在推进“数字单一市场”建设的进程中,欧盟开始在数字领域大量制定条例。[⑤] 在《2019—2024优先计划》(Political Guidelines for the Next European Commission 2019—2024)中,欧盟委员会提出了在数据使用、在线平台、网络安全和人工智

① Joanna Mazur/Renata Wloch, “Embedding Digital Economy: Fictitious Triple Movement in the European Union's Artificial Intelligence Act”, *Social & Legal Studies*, Vol. 33, No. 1, 2024, pp. 104 - 123, here p. 105.

② European Commission Communication, “A Digital Agenda for Europe”, COM (2010) 245 final, 2010 - 05 - 19, https://eur-lex. europa. eu/LexUriServ/LexUriServ. do? uri=COM:2010:0245:FIN:EN:PDF, 访问日期:2024 - 03 - 03。

③ European Commission Communication, “A Digital Single Market Strategy for Europe”, COM (2015) 192 final, 2015 - 05 - 05, p. 3, https://eur-lex. europa. eu/legal-content/EN/TXT/? uri=celex%3A52015DC0192, 访问日期:2024 - 02 - 04。

④ 金晶:《数字时代经典合同法的力量——以欧盟数字单一市场政策为背景》,载《欧洲研究》,2017年第6期,第65 - 89页,这里第66页。

⑤ 欧盟法律文件按照对成员国约束力由强至弱依次分为条例(regulation)、指令(directive)、决议(decision)和推荐建议(recommendations and opinion)。其中,只有条例直接适用于所有成员国,无需成员国进行实施性立法,并且凌驾于各国国内规定之上。

能等科技领域的数字立法蓝图，包括《人工智能法》、《数字服务条例》(Digital Service Act)、《数字市场条例》(Digital Market Act)、《欧洲芯片条例》(European Chips Act)和《数据治理条例》(Data Governance Act)在内的多项立法议题。[①] 借助这些数字立法计划，欧盟既可以促进一体化建设、实现监管协同，又能够通过市场机制将法律规则与技术标准外部化，发挥“布鲁塞尔效应”，单边监管全球市场。[②]

作为欧洲数字立法版图的重要组成部分，该法自酝酿之初便被寄予多样愿景，立法不仅是为应对人工智能对民主、法治和自由秩序的冲击和威胁，也旨在防止因监管差异性导致市场碎片化，确立欧盟在关键领域的技术主权，确保欧盟能够在全球数字竞争中胜出。欧洲数字战略的目标定位，决定了该法日后的立法模式与价值取舍。

(二)《值得信赖的人工智能的伦理准则》等软法规范的确立

由于人工智能领域存在大量立法空白，欧盟立法者首先转向软法规范，确立人工智能伦理准则，增加社会讨论度、凝聚社会共识，为后续立法奠定基础。2017年，欧洲议会法律事务委员提出了一份与民用机器人相关的立法动议，建议欧盟委员会制定一套有关机器人的民事规则。[③] 随后，欧盟委员会出台一份决议，指出有必要“充分考虑人工智能在法律和道德层面的影响”。[④] 决议在伦理部分以“机器人三定律”[⑤]为开端，强调应当建立一个符合基本伦理的人工智能监管框架，并在附件中具体阐述了人工智能的四条伦理原则——善意原则、不伤害原则、自主性原则和公正原则。

---

① 更多信息，见欧盟委员会官方网站，https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en，访问日期：2024－03－03。

② 欧盟通过立法或者技术标准单边监管全球市场的案例屡见不鲜，其中最典型的例子莫过于2022年欧盟出台规定，要求欧盟境内智能手机、平板电脑等设备统一采用USB Type-C接口，苹果公司随后即在其推出的iPhone15系列中放弃使用了11年的闪存接口，转而采用Type-C接口。

③ 根据欧盟立法架构，欧盟委员会享有提交立法议案的专属权力，欧洲议会和欧盟理事会则可以提出立法动议，建议欧盟委员会出台立法议案，并主要负责条例的修订与批准工作，因此首先由欧盟在议会提出立法动议。Committee on Legal Affairs, “Report with Recommendations to the Commission on Civil Law Rules on Robotics”, 2015/2013(INL), 2017－06－27, https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html, 访问日期：2024－02－28。

④ European Parliament, “Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics”, 2015/2103(INL), 2018/C 252/25, 2018－06－18, pp. 239－257, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&from=IT, 访问日期：2024－02－28。

⑤ 科幻小说家艾萨克·阿西莫夫(Isaac Asimov)在其1942年发表的短篇小说《环舞》(Runaround)中提出了一套虚构的机器人准则，分别是：1.机器人不得伤害人，也不得因不作为而使人受到伤害；2.机器人必须服从人的命令，除非与第一定律相冲突；3.机器人必须保护自己，只要不与第一和第二定律相冲突。该定律也被称为“机器人三定律”或“阿西莫夫定律”。

建立高度抽象的人工智能伦理原则很简单，而将其转化为可操控、可执行的行为要求则具有一定挑战性。为此，欧盟委员会组建了一支高水准专家团队（AI High-Level Expert Group，HLEG），该团队于2019年4月制定出《值得信赖的人工智能的伦理准则》（Ethics Guidelines for Trustworthy AI and Policy and Investment Recommendations，以下简称《伦理准则》）。该准则不具有强制性，而是围绕决议建立“可信赖”“以人为中心”人工智能的目标提出了具体实施框架和伦理要求，包括七项关键伦理要求、两条实施路径和一套评估清单。其中，伦理要求包括：（1）人类代理和监督；（2）技术稳健性和安全性；（3）隐私和数据治理；（4）透明度；（5）多样性、非歧视和公平；（6）社会和环境福祉；（7）可问责性。[①] 实施路径则包括技术性和非技术性两类：“技术性路径”是指在设计、开发和使用人工智能时采取特定的技术方法，使其能够符合伦理与法治要求，如要求算法具有可解释性或者通过技术测试与验证，以确保人工智能架构的合理性与合法性。[②] “非技术性路径”是指通过制定法规及标准、认证和独立审计等方式要求人工智能系统符合伦理要求。[③] 此外，《伦理准则》还附带发布了一份评估清单与基本权利影响评价表，用于指导企业在开发和设计人工智能系统前，通过自我评估的方式具体执行每项伦理要求，并研判系统对基本权利造成的影响情况。[④] 《伦理准则》只是一部软法规范，并不具有法律约束力，但其中的伦理要求最终被立法者吸纳借鉴，并融贯到欧盟《人工智能法》之中。

（三）向硬法的转变与《人工智能法》的最终出台

《伦理准则》等软法工具虽可影响企业的人工智能决策，引导人工智能系统朝着安全、可信和合法的方向发展，但这不意味着软法规范和自律倡议可以代替具备约束力和强制性的硬法规范。2019年12月，欧盟委员会新任主席乌尔苏拉·冯德莱恩（Ursula von der Leyen）在上任时的竞选发言中提出要打造“适应数字时代的欧洲”，并宣布将在任职的100天内，就人工智能对人类及伦理的影响提出一个统一的欧洲方案。[⑤] 2020年2月，欧盟委员会发布《人工智能白皮书》（White Paper

① High-Level Expert Group on Artificial Intelligence，“Ethics Guidelines for Trustworthy AI and Policy and Investment Recommendations”，2019－04－08，https://www.aepd.es/sites/default/files/2019－12/ai-ethics-guidelines.pdf，访问日期：2024－02－28。

② Nikos Th. Nikolinakos，*EU Policy and Legal Framework for Artificial Intelligence，Robotics and Related Technologies—The AI Act*，Cham：Springer，2023，p. 183.

③ 同上，pp. 183－184。

④ 同上，pp. 187－190。

⑤ Ursula von der Leyen，“A Union that Strives for More：My Agenda for Europe”，2019－10－09，p. 13，https://commission.europa.eu/document/download/063d44e9-04ed-4033-acf9-639ecb187e87_en?filename=political-guidelines-next-commission_en.pdf，访问日期：2024－02－28。

On Artificial Intelligence — A European Approach to Excellence and Trust)，详细地梳理了人工智能的潜在风险，[①]明确提出人工智能的法律框架应当“合比例”，并强调要以高风险领域的人工智能为监管核心，避免因过度监管而遏制创新。[②]该白皮书奠定了欧盟日后以风险为中心、注重高风险人工智能监管的立法格局。

随后，欧盟《人工智能法》正式进入立法轨道，历经 4 年最终出台。2020 年 6 月，欧洲议会成立专门委员会，致力于人工智能立法工作。该委员会主张形成欧洲共同的、整体的法律方案，并于同年 10 月通过决议，敦促欧盟委员会提出立法提案，以进一步明确在欧盟部署、开发和使用人工智能的伦理原则和法律义务。[③]2021 年 4 月 21 日，欧盟委员会提出了《人工智能法》立法草案，立法程序正式启动。此后，欧洲议会和欧盟理事会就该法进行了多轮修订及讨论。2023 年 6 月，欧洲议会以多数票通过的谈判授权草案，成为欧洲议会历史上讨论协商次数最多的文本之一，修正案多达 1000 多项。[④] 2023 年底，欧盟理事会、欧洲议会和欧盟委员会就立法文本达成三方共识，并于 2024 年 5 月 21 日最终通过《人工智能法》。

欧盟《人工智能法》规定了多重立法目标，具有明显的功能主义立法特征。该法序言部分指出，其立法目标包括：(1)改善欧洲内部市场运作；(2)发展以人为中心和值得信赖的人工智能应用；(3)保障《欧盟基本权利宪章》规定的包括健康、安全、民主与法治在内的基本权利；(4)防范人工智能风险；(5)支持创新。[⑤] 整体上看，该法包括了三种功能预设：一是市场整合功能，即促进欧盟数字市场建设的一体化，防止人工智能监管碎片化，实现监管协同。二是市场塑造功能，包括促进创新和防控风险两个面向，通过对高风险人工智能的重点监管将法律规范嵌入数字市场，并将人工智能风险限制在安全可控的范围内，同时又不至于过度遏制创新。三是基本权利保护功能，即保障安全、健康、自由、平等、公平、隐私等《欧盟基本权

① 这些风险包括侵犯个人隐私和数据保护、干预言论自由与政治自由、造成歧视和妨碍司法公正。

② European Commission Communication, “White Paper on Artificial Intelligence — A European Approach to Excellence and Trust”, COM (2020) 65 final, 2020 - 02 - 19, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/? uri=CELEX:52020DC0065, 访问日期：2024 - 02 - 28。

③ European Parliament, “Decision of 18 June 2020 on Setting Up a Special Committee on Artificial Intelligence in a Digital Age, and Defining Its Responsibilities, Numerical Strength and Term of Office”, 2020/2684(RSO), 2020 - 06 - 18, https://www.europarl.europa.eu/doceo/document/TA - 9 - 2020 - 0162_EN.html, 访问日期：2024 - 02 - 28。

④ 按照计划，欧洲议会原本应当在 2023 年 3 月就立法提案进行投票并达成立场，但由于立法者们就条例核心内容存在诸多争议，且 ChatGPT 等生成式人工智能技术为监管提出了新难题，因此条例出台进程一再推迟。参见徐路易、周芊屿、李子璇：《欧盟〈人工智能法案〉将落地》，载《财新周刊》，2024 年 2 月 19 日，https://weekly.caixin.com/2024 - 02 - 17/102166053.html? originReferrer=caixinweibo，访问日期：2024 - 02 - 20。

⑤ 《人工智能法》，序言 1。

利宪章》所规定的公民基本权利不受人工智能风险所侵害。[①] 但是，这三重相互平行的功能预设间存在难以调和的张力。例如，由欧盟立法设置监管“底线”，再由成员国在此基础上各自立法，可以为基本权利提供更高水平的保护，但这将阻碍统一市场建设；又如，强化技术监管与促进技术创新的目标左右互搏，无监管或弱监管环境也许更有助于激励创新。因此，我们有必要进一步探析，欧盟立法者是通过何种法律规制路径对存在内在张力的多重立法目标进行平衡的。

二、人工智能的三种法律规制路径

纵观欧盟人工智能政策的演进与立法讨论，我们可以将法律规制路径大致分为三种，分别是“基于原则”“基于权利”和“基于风险”的规制路径。[②] 诚然，三种路径并非截然不同，而是存在交叠重合，这样划分虽存在失准的风险，但仍有助于厘清不同路径选择的底层逻辑与立法考量。

（一）“基于原则”的规制路径

在欧盟人工智能立法进程中，最先出现的是“基于原则”的规制路径。“基于原则”的规制路径也被称为“原则治理”，旨在为人工智能发展设定基本、共通与关键性原则和发展目标，并要求企业以自我规制的方式尽力实现这些目标。[③] 原则和目标并不具备强制性和拘束力，而是有赖于企业自我实施，本质上属于软法治理模式。HLEG 2019年发布的《伦理准则》就采取了这种规制路径，并基于《欧盟基本权利宪章》中的基本权利体系提出了人工智能发展的四项基本原则，即尊重人类自主性、预防伤害、具有公平性和可解释性，分别与基本权利中的人的尊严、健康、平等和自由权相对应。除欧盟外，其他国家政府、非政府组织、公民社会、学术界以及私营部门也在过去几年中纷纷提出人工智能的发展原则。例如，美国《人工智能权利条例蓝图》(Blueprint of Artificial Intelligence Bill of Rights)提出人工智能需遵循安全有效、反算法歧视、数据隐私保护、通知解释、用户选择和退出自由五项原则；日本《人工智能应用指南》提出“以人为中心的人工智能社会”原则，并进一步提出人工智能社会发展的七项具体目标。据算法观察组织(Algorithm Watch)对全球人工智能伦理准则的梳理，截至2020年4月，全球共计出现超过160个伦理框架和道德准则。[④]

---

① Joanna Mazur/Renata Wloch, “Embedding Digital Economy: Fictitious Triple Movement in the European Union's Artificial Intelligence Act”, p. 111.

② Alessandro Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, The Hague: T. M. C. Asser Press, 2022, p. 141.

③ Ren Bin Lee Dixon, “A Principled Governance for Emerging AI Regimes: Lessons from China, the European Union, and the United States”, *AI and Ethics*, Vol. 3, No. 3, 2023, pp. 793 – 810, here p. 793.

④ Algorithm Watch, “AI Ethics Guidelines Global Inventory”, https://inventory.algorithmwatch.org，访问日期：2024 – 03 – 20。

这些原则极大地丰富了人工智能的软法治理框架，为人工智能发展注入了价值内核。

“基于原则”的人工智能规制路径不仅指宽泛、抽象的人工治理原则，还包括了一套完整原则的形成过程和具体、可操作的实施方法。亚历山德罗·曼特莱罗(Alessandro Mantelero)提出，人工智能治理原则有两种形成方法：一是以权利为中心的方法，即从基本权利体系中抽离出彼此独立的权利，然后分析人工智能会对这些权利产生何种影响，欧盟《伦理准则》中提出的四项原则就源自这种方法；二是以领域为中心的方法，即对人工智能的应用领域进行具体分析，例如医疗、教育、司法等，然后从这些纵向领域中各自提炼出共通、相似的要素，并形成领域性的治理原则。[①] 以领域为中心的方法强调在不同领域中提炼出差异化的人工智能治理原则，使其“既不脱离现有法律框架，也不停留在对每一项权利和自由的抽象理论的构想上”。[②]

该规制路径的优势在于灵活性和包容性，可以在具体应用场景中将高度抽象凝练的原则情景化，对于快速发展的技术表现出更强的适应性，具有未来面向。相反，过于具体的法律规则则可能在技术创新方面存在滞后、僵化和失败的风险。然而，“基于原则”的规制路径也存在显著的局限性：一是缺乏统一性。多方主体发布的大量原则构成了“原则市场”(principle market)，在这个市场中，企业可能选择性地遵从有利于自己的原则和伦理规范，从而出现“挑选原则”(principle forum)的情况。二是缺乏强制性。经验研究表明，无论是原则、道德准则还是伦理规范，对人工智能产品开发者决策的影响并不大，人工智能产品开发者们往往认为这些原则是“由技术界以外的机构强加的”，至于何时纳入伦理要求，则“主要出于营销策略的考虑”。[③] 三是缺乏明确性。如果仅按照松散、模糊的原则对人工智能相关活动进行监管，日后的执法和司法将面临诸多挑战。例如，由于人工智能治理原则通常由不确定法律概念构成，在没有充分量化标准的情况下，企业将很难根据这些原则开展合规工作。此外，诸如公平性、可解释性等人工智能治理原则也存在很大的解释空间，行政机关在根据这些原则进行监管或处罚时，也可能受到被监管对象的质疑，后者还可能启动司法程序进行抗辩，因此给执法带来更大挑战。

(二)“基于权利”的规制路径

相比“原则治理”而言，“基于权利”的路径在规制型立法中更为常见。从广义

① See Alessandro Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, pp. 142 - 144.

② 同上，p. 143。

③ T. Hagendorff, “The Ethics of AI Ethics: An Evaluation of Guidelines”, *Minds and Machines*, Vol. 30, 2020, pp. 99 - 120, here p. 99.

上看，凡是将基本权利保护放在首位、强调以人为本和以保护为中心的治理方法都可以视为“基于权利”（或“基于人权”）的规制路径。[①] 从狭义上看，“基于权利”的规制路径是指赋予公民或法人控制性权利、对权力形成制衡，并以此方式间接规制对权力的运用。例如，在个人信息保护中，信息主体被赋予了知情同意权、查阅权、请求更正权和删除权，以便对公共部门、商业部门等信息处理者的权力进行制衡，使后者透明、负责任地处理个人信息，并维护受影响个体的权益。[②] 本文仅在狭义层面使用此概念。

权利路径最初并未被纳入欧盟立法者的考量之中，而是在立法辩论过程中，凭借公民社会、非政府组织和学界不断的建议和呼吁，才逐步落实在《人工智能法》最终文本中。第一，基于公民社会对权利保障的呼声，立法者在原本高/低两类人工智能风险的基础上增加了“被禁止的风险”这一类型，提高了监管强度。欧盟委员会最初发布的《人工智能白皮书》仅将人工智能划分为高/低两类风险，并不包括被禁止的类别。对此，有批判者认为，这代表在立法价值排序上，欧盟委员会将促进创新放在了保护基本权利之前，并建议应当禁止在公共场所使用构成大规模监控的面部识别系统、对跨性别者构成歧视的性别识别系统等严重侵害基本权利的人工智能系统。[③] 非政府组织“欧洲数字权利”（European Digital Rights）也指出，高/低风险的分类方法无法充分预防、减轻和救济基本权利侵权行为，建议应当在必要时通过侵权诉讼的方式禁止生物识别处理行为，并要求高风险人工智能系统的提供者对该体系进行强制性的基本权利影响评估。[④] 对此，在后续发布的立法草案中，欧盟委员会完善了风险类型，设置了禁止风险、高风险、有限风险和低风险四种风险等级。第二，为应对学界、公民社会和政府组织对该法权利保障不足的批判，立法者在修改过程中逐步增加了赋权条款。起初，立法者虽然关注人工智能对公民基本权利的影响，但并未直接规定赋权条款。例如，2020年底，欧盟理事会下设的人工智能特设委员会出具了一份立法可行性研究报告，报告详细地梳理了人工智能所涉的公民基本权利，包括人格权、生命权、健康权；与人工智能系统互动时的

① Scottish Human Right Commission, “What is Human Right Based Approach?”, https://careaboutrights.scottishhumanrights.com/whatisahumanrightsbasedapproach.html, 访问日期：2024－03－20。

② 赵鹏：《“基于风险”的个人信息保护？》，载《法学评论》，2023年第4期，第123－136页，这里第124－125页。

③ Fanny Hidvegi Massé Daniel Leufer, Estelle, “The EU Should Regulate AI on the Basis of Rights, Not Risks”, 2023－01－13, https://www.accessnow.org/eu-regulation-ai-risk-based-approach/，访问日期：2024－03－20。

④ European Digital Rights, “Recommendations for a Fundamental Rights-Based Artificial Intelligence Regulation: Addressing Collective Harms, Democratic Oversight and Impermissable Use”, 2020－06－20, https://edri.org/wp—content/uploads/2020/06/AI_EDRiRecommendations.pdf, 访问日期：2024－03－20。

知情权；拒绝权和退出权；要求非歧视与平等对待权；对于人工智能决策的及时告知权和要求解释权，并提出了相对应的法律义务。[①] 然而，2021 年欧盟委员会发布的立法草案仅在序言部分强调保障公民基本权利，未在正文规定任何对个体权利的赋权条款，也未承认受损害用户的赔偿请求权。对此，有批评者认为该法的法律框架完全是从企业角度出发进行的设计，忽视了公民的权利保障。[②] 直至 2024 年 2 月，欧洲议会、欧盟理事会和欧盟委员会才在其发布的三方商定版本中新增了救济专章，赋予了公民向市场监督管理机关投诉的权利（第 68a 条，最终版中为第 85 条）。随后，在 2024 年 4 月发布的最终版中，该法再次新增了针对高风险人工智能系统决策要求解释的权利（第 86 条）。

而在数字立法领域，“基于权利”的规制路径也面临着质疑与挑战。以立法形式直接赋予公民针对不法行为的诉权或司法救济请求权将面临两方面挑战：一是制约不足。相较于人工智能提供者而言，公民个体在举证能力和诉讼能力方面都处于弱势地位，私人权利和数字权力间呈现强大势差，这可能导致对权力的约束不足。二是过度干扰。采取私人诉讼的方式还可能走向另一极端，即出现恶意诉讼或者滥诉现象，从而扰乱科技企业正常的生产经营活动，阻碍市场发展和科技创新。实际上，在个人信息保护领域，就有学者分析了权利路径存在的缺陷，指出从赋权意义上理解个人信息控制可能导致个人信息的私权化，与个人信息的公共性相悖，不利于信息的流通利用。[③]

（三）“基于风险”的规制路径

“基于风险”的规制路径是欧盟《人工智能法》中最显著的特征，该法在序言部分明确说明采取“基于风险”的规制路径。[④] 其核心在于以前所未有的力度强调防范风险的重要性，并将风险作为触发或缓和法律规制的主要因素。在欧盟立法史上，“基于风险”的规制路径并非新鲜事物。早在 1995 年的《数据保护指令》（Data Protection Directive）中便蕴含了这种思路，该指令规定了数据保护措施与风险等

① Ad hoc Committee on Artificial Intelligence, “Feasibility Study”, Section 7, 2020 - 11 - 17, https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da, 访问日期：2024 - 03 - 20。

② Costanza Alfieri/Francesca Caroccia/Paola Inverardi, “AI Act and Individual Rights: A Juridical and Technical Perspective”, in Proceedings of the Workshop on Imagining the AI Landscape after the AI Act, Vol. 3221, 2022, pp. 43 - 55, here p. 43.

③ 参见高富平：《个人信息保护：从个人控制到社会控制》，载《法学研究》，2018 年第 3 期，第 84 - 101 页，这里第 98 - 99 页。其他相类似观点与讨论，参见周汉华：《个人信息保护的法律定位》，载《法商研究》，2020 年第 3 期，第 44 - 56 页；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》，载《法学研究》，2022 年第 5 期，第 3 - 18 页。

④ 《人工智能法》序言 26 指出，“应当遵循‘基于风险’的方法，为人工智能系统引入一套相称且有效的约束性规则”。

级相适应(第17条第1款)和数据保护官制度(第20条),是"基于风险"的规制路径的雏形。[①] 2016年出台GDPR进一步发展了这种方法,将数据处理保护措施与风险合乎比例确定为一般原则(第26条),并规定了个人数据保护影响评估制度(第33条)。在GDPR中,风险具有双重功能:对于监管者而言,风险是用于实施监督或者控制的标准和触发器;对于数据处理者而言,风险是其用以校准自身法律义务的参考点。[②] 在人工智能治理中,"基于风险"的规制路径是指根据不同的人工智能类型和风险等级,采取相应的治理措施和方式,实现以风险为中心的治理模式。例如,欧盟《人工智能法》根据"风险金字塔"理论将人工智能风险划分为不可接受风险、高风险、有限风险和低风险四个等级,并围绕不同风险等级提出相应的禁令、风险管理、风险合规与合规评定要求。

该路径有四个主要特征:一是在规制重心上,通过强化企业负责性而非个体控制性权利,要求企业通过尽力合规的方式满足风险管理和风险预防的要求,从源头防止风险产生。其核心理念在于,相较于政府或个人而言,企业作为人工智能技术的拥有者,最具备风险防控的能力和资源,即"让最有能力控制和管理风险的一方负责"。二是在规制模式上,强调企业的元规制或自我规制,而非"命令-控制"式政府规制,在国家通过立法建立风险合规框架后,由企业主动履行风险合规义务。三是在风险控制方式上,"基于风险"的规制路径更侧重于从社会层面进行整体性的风险防范,而非通过公民诉讼的方式对风险进行个体控制。四是在监管措施上,强调监管手段的可伸缩与合比例性,通过灵活且富有弹性的监管措施,平衡预防风险与促进创新的关系。[③]

"基于风险"的规制路径的提出和发展,体现了欧盟立法者自GDPR以来在个人信息保护、人工智能监管等数字治理中法律规制的范式转型,具体可以分为两方面:一是在实践方面,向基于风险的执法与合规转变;二是在监管方面,向风险监管转变。[④] 这一转型有深刻的现实背景。第一,在数字化、信息化时代,风险的类型

① European Parliament and the Council of the European Union, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with regard to the Processing of Personal Data and on the Free Movement of Such Data", 1995-11-21, Art. 17, 20, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31995L0046, 访问日期:2024-04-22。

② Ronald Leenes et al. (eds.), *Data Protection and Privacy: The Age of Intelligent Machines*, Oxford: Hart Publishing, 2017, p. 40.

③ 有关"基于风险"路径特征的更多讨论,参见张涛:《人工智能治理中"基于风险的方法":理论、实践与反思》,载《华中科技大学学报(社会科学版)》,2024年第2期,第66-77页;赵鹏:《"基于风险"的个人信息保护?》,载《法学评论》,2023年第4期,第123-136页。

④ Milda Macenaite, "The 'Riskification' of European Data Protection Law through a Two-Fold Shift", *European Journal of Risk Regulation*, Vol. 8, No. 3, 2017, pp. 506-540, here p. 506.

和特征发生了变化。过去，法律通常是将社会中已经常见的社会关系通过规则化的方式予以固定总结，调整对象是社会生活中存在已久的事物，其法律风险和法律关系是已知的、确定的。而现在，人工智能、数据、算法所带来的风险具有快速变化性和不确定性，所引发的风险不再是单一的、个体性的，而是系统的、社会性的，因此需要国家进行整体规制，从社会层面预防风险的产生和爆发。第二，在面对人工智能侵权时，通过诉讼的方式进行私人执法或采取其他风险的个人控制手段存在局限性。一方面，司法救济成本高昂、程序进展缓慢，并非所有的私人主体都有能力提起诉讼。私人主体可能要么放弃起诉，要么仅就产业价值链中较薄弱的环节提起诉讼。另一方面，即便私人主体有能力提起诉讼，也可能因技术的复杂性而难以成功举证。相比而言，“基于风险”的规制路径通过强调企业事先责任使最具技术能力的人控制风险并将风险成本充分内部化。[①] 第三，在人工智能时代，受保护的法益也产生了变化。侵权之诉中需要证明受保护的法益与侵权行为的相关性。然而，区别于一般的侵害健康权、生命权的产品侵权情形，人工智能产品更多的在于对自由权的侵害，例如因人脸识别造成的大规模监控现象，而这种侵害通常是非常微弱、难以觉察的，以至于很难评估甚至根本无法评估，因此许多立法提案更多地关注技术本身，而非个人受保护的法律利益。此外，人工智能对基本权利的威胁程度需要依据具体应用环境而定，不同人机交互环境的风险特征差异明显。因此，相较于“基于权利”的规制路径而言，“基于风险”的规制路径是更恰当、更符合人工智能技术特征与风险特征的规制路径。[②]

然而，“基于风险”的规制路径也具有局限性，可能造成个体权利保护的缺失，尤其是在仅采取风险路径而忽视权利路径时。在此，可以将欧盟《人工智能法》与GDPR进行对比：GDPR在第三章第12—23条中清晰地列出了数据主体的权利，并在第77—84条中明确了数据主体可以就侵权行为获得行政、司法救济和损害赔偿的权利。反观欧盟《人工智能法》，虽然立法者在一定程度上吸纳了批判意见，在修改过程中增设了两条赋权条款，但仍未规定公民的司法救济和损害赔偿权。对此，有观点指出，“事先安全应当与事后损害相脱钩”，仅采取“基于风险”的规制路径，从社会角度考虑如何预防或减少人工智能的整体社会危害，将不可避免地忽视人工智能给个体权利造成的具体损害，唯有将事先责任条款与损害赔偿义务条款

① Andrea Bertolini, “Artificial Intelligence and Civil Liability”, study requested by the European Parliament's Committee on Legal Affairs, 2020－06, p. 100, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf, 访问日期：2024－03－20。

② Hannah Ruschemeier, “AI as a Challenge for Legal Regulation — The Scope of Application of the Artificial Intelligence Act Proposal”, *European Research Area Forum*, Vol. 23, No. 3, 2023, pp. 361－376, here p. 364.

相结合，才能打造出一个真正以人为中心、可信赖的人工智能系统。①

“基于权利”和“基于风险”经常被学界作为一组相对应的概念进行讨论，二者呈现出法律规制的不同侧重点。“基于权利”的规制路径更侧重于对公民和法人主体的赋权，个体权利的存在对被规制主体行为起到威慑作用，使得后者不得不在行动时考虑这种保护性权利。在法律执行方面，私人主体通过提起诉讼的方式维护自身权益，同时这类诉讼也具有“私人执法”的效果，是对公共部门执法的有效补充。“基于风险”的规制路径则更多地着眼于企业而非个体层面，例如对企业提出风险评估和合规义务，且这种义务更多地体现为一种程序性而非实体性义务，并在企业违反义务时通过行政手段而非司法手段进行威慑（如巨额行政罚款）。从本质上看，“基于权利”的规制路径体现了对潜在风险的个人控制，“基于风险”的规制路径则是一种对风险的社会控制。以个人信息保护领域为例，国家立法承认公民对个人数据享有个人自治和自决的法益，并将个人数据保护权作为一项宪法层面的基本权利，赋予信息主体对个人数据的自决权和控制权，以制约信息持有者的个人信息采集行为，体现了一种对个人信息的个人控制；与之相对应，国家立法着眼于个人信息的公共性和社会性，保护个人信息的责任主体由以个人为主转向以社会为主，法律规范重心从收集行为转向使用行为，保护模式由个人维权转变为行政监管，体现了一种个人信息的社会控制。② 不过，将“基于权利”和“基于风险”二元对立的化约主义思维也容易引发误解。二者虽然强调保护的侧重点不同，但并不互相排斥，甚至可以同时出现在一部立法之中。例如，即使是在采取“基于风险”的规制路径的GDPR中，也存在赋权条款，后者为数据主体构建了一系列保护性权利，包括数据可携权、删除权（被遗忘权）、访问权、限制处理权和拒绝权等。③ 事实上，在GDPR起草过程中，欧盟数据保护第29工作组对“基于风险”的规制路径进行了特别澄清，表示仅采取“基于风险”的规制路径无法充分支撑起欧盟数据保护的法律框架，强调无论在数据处理过程中产生的风险程度如何，法律都应该赋予数据主体权利。④

① Andrea Bertolini, “Artificial Intelligence and Civil Liability”, p. 100.

② 参见高富平：《个人信息保护：从个人控制到社会控制》，载《法学研究》，2018年第3期，第84－101页，这里第99－100页。

③ 参见金晶：《欧盟〈一般数据保护条例〉：演进、要点与疑义》，载《欧洲研究》，2018年第4期，第1－26页，这里第13－17页。

④ 29 Data Protection Working Party, “Statement on the Role of a Risk-Based Approach in Data Protection Legal Frameworks”, WP 2018 (2014), 2014－05－30, p. 2, https://ec.europa.eu/justice/article—29/documentation/opinion—recommendation/files/2014/wp218_en.pdf, 访问日期：2024－04－01。

三、风险路径下《人工智能法》的立法要点与核心制度

为调和多重法益间的内在张力，欧盟立法者最终选择采取“基于风险”的规制路径，通过强化企业负责性而非赋权公民个人的方式平衡监管与创新之间的矛盾。在以风险为中心的立法格局下，《人工智能法》对人工智能进行了定义，设计了禁止规则、合规规则和豁免规则三类人工智能监管规则，以期实现监管的可伸缩性和合比例性。

（一）人工智能的定义

定义问题淋漓尽致地展现了技术的动态性、迭代性与法律的一般性、理论性之间的冲突与张力。对被规制主体进行界定决定了规制手段的适用范围，定义过宽会超越规制目标，定义过窄则无法实现保护和管理的目的。因此，定义必须具有准确性、全面性和实用性，使法律具有确定性、可预见性和可适用性。[①] 此外，法律定义还应当具有弹性、包容性和发展性，能够为大量案例创造抽象的一般规范，从而实现法律概念的存储器功能。但作为一项跨领域技术，人工智能并不存在一个固定且通用的定义，“人工智能”一词不仅具有高度模糊性，而且还会随着时间变化和新技术涌现而改变。[②] 人工智能的不透明性、复杂性、算法偏差等特征与法律要求的精确性、透明性、可预期相抵牾。

因此，在立法进程中，人工智能的定义条款几经更迭。在立法草案中，人工智能系统被定义为“使用附件一中列出的一种或多种技术方法开发的软件，它可以根据一组人为设定的目标生成内容、预测、建议或决定等影响其与交互环境的输出”。[③] 该定义存在两个问题：第一，将人工智能系统定位为“软件”，忽视了人工智能的硬件部分，因此立法草案也被学者戏称为“软件条例”。[④] 此外，采取“软件”的定义方式可能使立法产生抑制创新的效果，使企业停止对其硬件产品进行升级，通过“去智能化”的方式规避法律要求。第二，要求人工智能系统基于“人为设定的目标”行动，忽视了其自主性和自我学习能力，并且预设人工智能的歧视、偏见行为背后可能存在一个“人为目的”，忽略了系统自我偏离的可能性。例如，人脸识别系统无法识别深色皮肤女性，这可能是因为其所学习的数据集样

① Hannah Ruschemeier, “AI as a Challenge for Legal Regulation — The Scope of Application of the artificial Intelligence Act Proposal”, p. 366.

② Jonas Schuett, “Defining the Scope of AI Regulations”, *Law, Innovation and Technology*, Vol. 15, No. 1, 2023, pp. 60 – 82, here p. 68.

③ European Commission, “Proposal for Artificial Intelligence Act”, article 3 “definition”.

④ 同注①，p. 361。

本量有限，无法充分地代表整个社会人群分布，但并不一定意味着系统开发者具有种族歧视或者偏见。①

有鉴于此，终稿对“人工智能”这一核心定义进行了一定程度变更，将人工智能系统定义为“为了明确或隐含的目标，从输入信息推断如何生成诸如预测、内容、建议或可能影响物理或虚拟环境的决策等输出信息的一种基于机器的系统”。② 新的定义有两处优化：第一，“基于机器的系统”一词同时涵盖了软件和硬件，修正了原先的定义偏差；第二，以“推断”(infers)能力将人工智能系统与其他自动化系统区别开来。立法者解释，“人工智能系统的关键性特征是它们的推断能力，通过从机器学习的方法或基于逻辑的方法，从待解决任务中的知识编码或符号中进行推断，这种推断能力超越了基本的数据处理，使该系统可以进行学习、推理和建模”。③ 该解释将两种人工智能系统，即机器学习系统和基于符号或知识的系统(也称专家系统)都纳入其中，拓展了原定义的使用范围。

(二)禁止、合规和豁免规则

对人工智能的监管规则是《人工智能法》的主体内容，这些监管规则可以大致分为三类：1.针对不可接受风险的禁止规则；2.针对高风险、低风险和有限风险的合规规则；3.针对特殊情况或用以鼓励创新的豁免规则，以下分而述之。

1. 针对不可接受风险的禁止规则

该法第5条规定了四类被禁止的人工智能系统，其中三类为完全禁止，一类为相对禁止。完全禁止的人工智能系统包括操纵系统、剥削系统与社会评分系统。(1)操纵系统，指利用“潜意识技术”④对个体意识进行操纵或欺骗，从而扭曲或诱导个人或群体决策，并损害、颠覆人类自主决策与自由选择的人工智能系统。⑤ 例如，脑机界面或者虚拟现实就可能因为对人脑的过度刺激而实质性扭曲个体行为。(2)剥削系统，指利用个体或群体因年龄、残疾或特定社会或经济情况的弱点，实质性扭曲其行为并对其造成严重损害的系统，但合法运用于医疗领域中的系统除外。⑥ (3)

① See Eduard Fosch Villaronga/Adam Poulsen, “Diversity and Inclusion in Artificial Intelligence”, in Bart Custers/Eduard Fosch Villaronga (eds.), *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, The Hague: T. M. C. Asser Press, 2022, pp. 109 - 134, here p. 109.

② 《人工智能法》，第3条第1款。

③ 《人工智能法》，序言12。

④ 潜意识技术(subconscious technique)是指使用高于或低于人类有意识感知阈值的感官刺激(如图像、文本或声音)的技术。参见《人工智能法》，第33条33b款。

⑤ 《人工智能法》，第5条第1款a项。

⑥ 《人工智能法》，第5条第1款b项。该系统包括一个重要子系统，即生物识别分类系统，指基于生物识别数据(例如脸部数据或者指纹数据)来推断种族、政治观点、工会成员身份、宗教或哲学信仰、性生活和性取向的系统。

社会评分系统，指根据个人或群体的社会行为来推断或预测个人人格特征，并对其进行评估和分类的系统。[①]

相对禁止的人工智能系统是指在公众场所出于执法目的而使用的“实时”远程生物识别系统。它将极大地影响公民隐私和私人生活，使人产生始终受到监视的感觉，并且会间接地妨碍集会自由或者其他基本权利的行使。[②] 与前三类禁令不同，该禁令仅仅禁止“使用”，而不禁止“投入市场”，这意味着欧盟供应商可以将此系统出售给欧盟之外的国家。此外，在立法谈判过程中，欧盟多次放宽对实时远程生物识别系统例外情况的限制，因此有批评者认为其并非真正的禁令。[③]

2. 针对高风险和剩余风险的合规规则

对于高风险和剩余风险（包括有限风险和低风险）人工智能系统，该法主要采取事前合规规则，重点关注高风险人工智能系统。

（1）高风险系统合规。高风险人工智能系统包括两类，一是作为产品或系统安全组件的人工智能系统（产品类）；二是独立的人工智能系统（系统类）。[④] 合规过程贯彻人工智能的整个生命周期，后者包括四个阶段：设计研发阶段、合格评估阶段（上市前）、系统注册阶段（上市前）和监测阶段（上市后）。在设计研发阶段，系统须满足数据治理、编制技术文档、记录保存的要求；在评估阶段，人工智能提供者须对系统进行合格评定；在注册阶段，人工智能提供者须将人工智能系统注册到统一的欧盟数据库中，并由供应商为系统制作合规声明、粘贴欧盟统一的安全认证标志（European conformity，CE）；在监测阶段，供应商须建立售后监测系统，并将故障及时通知监管机构。[⑤]

在高风险人工智能系统合规过程中，最关键的是要符合欧盟《人工智能法》第9条“风险管理系统”的要求。第9条是该法最重要的条款之一，由“风险管理过程”和“检测程序”两部分组成。风险管理过程包括风险识别、风险评估和风险应对

① 《人工智能法》，第5条第1款c项。典型的社会评分系统包括通过对自然人画像来评估和预测其潜在刑事犯罪可能性的系统，该系统有违无罪推定的刑事法律原则，也对个人造成严重的不平等对待。

② 《人工智能法》，序言17。该禁令存在三条例外情况，分别是：（1）为寻找特定犯罪受害者，例如搜寻失踪人员；（2）自然人生命或人身安全遭受特大威胁或恐怖袭击时；（3）为确定附件二a所述刑事犯罪嫌疑人位置或身份，前提是这些刑事犯罪会在有关成员国受到至少四年以上的监禁或者拘留判决。

③ Nikos Th. Nikolinakos, *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies — the AI Act*, p. 387.

④ 产品或者系统是否属于高风险等级由条例附件三所确定，其规定了八大高风险领域，包括：欧盟或成员国立法允许的生物识别技术领域；关键基础设施领域；教育和职业培训领域；就业领域；基本公共/私人服务与福利领域；执法领域；移民领域；司法领域。参见《人工智能法》附件三。

⑤ 参见曾雄、梁正、张辉：《欧盟人工智能的规制路径及其对我国的启示——以〈人工智能法案〉为分析对象》，载《电子政务》，2022年第9期，第63－72页，这里第66页。

(也称“风险处理”)三个步骤。第一步,风险识别,即识别并分析已知的风险或可预见的风险。[①] 第二步,风险评估,即对危害发生的概率和严重程度进行预估。[②] 第三步,风险应对或风险处理,即为减少已识别的风险而采取的行动。虽然识别、评估和处理风险三步骤是依次提出的,但风险管理系统本身就是一个不断迭代、持续和反复的系统,该过程需要不断重复,直到剩余风险减少到可接受水平。若高风险人工智能系统反复迭代后仍然无法达到可接受水平,则应当停止对其进行开发或部署。[③] 违反合规规则的高风险人工智能系统可能面临行政、民事和刑事责任。该法本身仅规定了行政责任,[④]民事或刑事责任则需视个案情况而定。例如,若合同双方将遵循该法第9条的规定作为合同附随义务,则可能产生民事责任;或者当欧盟成员国通过国内立法的方式将未遵循该法第9条规定引发的侵害定义为因疏忽大意构成的犯罪时,则可能产生刑事责任。

(2)剩余风险人工智能系统。有限风险和最小风险被统称为“剩余风险”。在条例的监管框架下,这类人工智能系统享有较高的自由度,无需在投入市场前进行认证或履行特定的审查、报告、监督、留存记录等义务。尽管如此,各行业仍可以建立自愿遵循的行为规范或道德准则来规范剩余风险类型的人工智能系统。也可以说,对于此风险类型而言,该法采取了一种“基于原则”的监管路径,强调自愿式的软法规制。

3. 针对监管沙盒与特殊豁免的豁免规则

为了最大限度地激励并保护创新,《人工智能法》还规定了豁免规则(也称“例外规则”),包括监管沙盒和特殊豁免。

(1)监管沙盒(regulatory sandbox)。[⑤] 监管沙盒最早被应用在金融监管领域,

① 如果风险在过去发生过或者未来肯定会发生,则风险就是“已知的”;如果风险尚未发生,但是已经可以识别,则该风险是“可预见的”。至于需要投入多少精力来识别“可预见的风险”,则需要遵循以下经验法则:风险的潜在影响越大,则需要投入预测和识别的经历越多。See Jonas Schuett, “Risk Management in the Artificial Intelligence Act”, *European Journal of Risk Regulation*, 2023, pp. 1 - 19, here pp. 9 - 10.

② 欧盟《人工智能法》并未直接定义“风险评估”,该定义来自ISO/IEC23894国际标准第三条。ISO/IEC Guide 51:2014 Safety Aspects — Guidelines for Their Inclusion in Standards, https://www.iso.org/standard/53940.html,访问日期:2024 - 04 - 01。此外,条例对被评估风险的范围进行了限制(第9条第2款b项),只涵盖了使用者“按照预期目的”和“在可合理预见的滥用条件下”使用高风险人工智能系统时可能出现的风险,不包括使用者未按照预期目的或者以人工智能系统提供者不可预见的方式滥用而产生的风险。

③ Jonas Schuett, “Risk Management in the Artificial Intelligence Act”, p. 12.

④ 对违反禁止人工智能系统有关规定的公司,欧盟将对其处以最高1500万欧元罚款,对于违反高风险人工智能系统有关规定的公司,欧盟将对其处以最高75万欧元的罚款。参见《人工智能法》第100条第2款、第3款。

⑤ 《人工智能法》,第57条。

后逐渐扩展应用于数据保护、航空等监管领域中。[①] 从本质上看，监管沙盒是一种监管工具或者监管过程，其核心是为颠覆式技术创新提供一个可控的测试场所或实验环境。在这个试验场中，企业可以通过测试减少潜在风险，监管部门也能在发生偏离时及时介入，不至于引发系统性风险。[②] 相比传统的监管工具，沙盒监管具有许多优势：首先，这种监管方式更为灵活、动态、具有包容性，能够鼓励技术创新。一方面，采取监管沙盒本身就体现了监管者促进和激励创新的意愿；另一方面，企业也能够在沙盒的范围内测试新技术，而免于行政制裁的威慑。其次，监管沙盒还能够促进规制学习。过往经验表明，在沙盒监管过程中，参与者和主管机构都能够提升各自对技术风险的认知水平和洞察力，从而使监管部门更好地履行其职责。最后，监管沙盒还能通过缩短新技术进入市场的时间，降低交易成本。[③] 但监管沙盒能否在人工智能领域发挥效用，还有待进一步观察。对此，有学者提出了两点顾虑：一是迄今为止，监管沙盒的目的并非测试技术，而是固定技术创新；二是人工智能作为一项跨领域应用，涉及多个主管部门，不同部门间的协调沟通也会为设定监管沙盒计划增加阻碍。[④]

(2)特殊豁免。除了监管沙盒，《人工智能法》还规定了四类豁免规则，分别适用于国家安全、军事国防、科学研究和出于开发目的的人工智能及部分开源的人工智能。其中值得关注的是后两类豁免，即专为科学研究与开发目的的人工智能系统及开源模型[⑤]的豁免。对开源模型进行豁免是基于其对研究、创新和竞争的积极影响，开源豁免属于一种反向豁免，即在所有风险等级的人工智能系统中，开源模型都可以享受豁免红利。

(三)对通用人工智能模型的规制

对“通用人工智能”(General Purpose Artificial Intelligence，GPAI)和“基础模型”的法律规制，亦构成欧盟《人工智能法》的一个重要维度。理论层面，通用人工智能与专用人工智能是一组对应概念，通用人工智能指用来解决多样、复杂、广泛的任务，并且“在各种任务中都等于或者超过人类智能”的一种未来形式的人工智能；专用人工智能指用来解决特定问题、完成特定任务的人工智能系统，是当前主

① Katerina Yordanova/Natalie Bertels，“Regulating AI：Challenges and the Way Forward Through Regulatory Sandboxes”，in Henrique Sousa Antunes et al.，*Multidisciplinary Perspectives on Artificial Intelligence and the Law*，Switzerland：Springer，2024，pp. 441－456，here p. 446.

② 同上，p. 447。

③ 同上，p. 449。

④ 同上，p. 450。

⑤ 根据定义，“开源模型”指“用户可以自由地访问、使用、修改并再次分析其改进版本的公开共享的模型”。参见欧盟《人工智能法》，序言 102。

要的人工智能形式。[①] 通用人工智能在实践中尚未出现，欧盟《人工智能法》中主要讨论的是技术层面的通用人工智能，即“通用人工智能模型”，指具有通用性特点，能够胜任广泛任务并在大数据基础上进行训练且具有自我学习能力的模型。该模型具有以下特征：一是具有庞大的数据训练量，二是可以执行非特定任务。其中，生成式人工智能就是通用人工智能模型的典例，以ChatGPT为例，其具有多样应用场景，如简历筛选、代码编写、论文写作等。对于通用人工智能而言，其风险取决于终端用户使用系统的方式，而非由系统供应商预先决定，这就可能使得基于预设用途进行风险分类的方式失效。[②] 因此，通用人工智能模型的兴起给欧盟《人工智能法》“基于风险”的规制路径带来了新的挑战。

为应对这一挑战，在立法讨论过程中，欧盟立法者增加了对通用人工智能模型的界定和法律规制方式，采用了一种层加式的风险分类方法，将通用人工智能模型分为开发许可型（open licensed）、标准型（standard）和系统性风险型（systematic risk），分别对应低、中、高三类风险类型。其中开发许可型GPAI具有可公开访问的参数和体系结构，是一种开源模型，提供者仅需满足技术文档的要求；标准型GPAI则需提供详细的技术和信息文档，让系统的下游用户了解系统的功能、不足、知识产权情况和训练数据透明度；系统性风险型GPAI则必须履行高风险类型系统的合规义务，包括风险管理、测试、向人工智能办公室报告以及提供网络安全保护等。[③]

四、欧盟《人工智能法》的反思与启示

欧盟《人工智能法》开人工智能领域法律规制之先河，填补了法律框架上的空白，为后来者提供了丰富的立法理论与经验素材。在我国，人工智能立法也成为未来发展趋势。国务院于2023年和2024年两度将“人工智能法草案”列入《立法工作计划》，我国学界也积极参与立法讨论，相关研究机构发布了《中华人民共和国人工智能法（学者建议稿）》《人工智能示范法2.0（专家建议稿）》等多部立法建议稿。下文将对欧盟《人工智能法》进行评述与反思，并在此基础上讨论其对我国人工智能立法的启示与借鉴意义。

① Tobias Mahler, “Regulating Artificial General Intelligence (AGI)”, in Bart Custers/Eduard Fosch Villaronga (eds.), *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, The Hague: T. M. C. Asser Press, 2022, pp. 521 - 540, here p. 521.

② Natali Helberger/Nicholas Diakopoulos, “ChatGPT and the AI Act”, *Internet Policy Review*, Vol. 12, No. 1, 2023, pp. 2 - 6, here p. 2.

③ Claudio Novelli et al., “Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity”, in the proceeding of SSRN Electronic Journal, 2024, p. 3, https://arxiv.org/abs/2401.07348, 访问日期：2024 - 04 - 01。

（一）对欧盟《人工智能法》的反思

立论之处，必有反思。自欧盟立法进程初始，商榷之声便不绝于耳。欧盟《人工智能法》是否代表了一种最佳立法方案，仍有待进一步考察和辨析。下文从该法的立法思维、立法模式、规制路径、整体构造和立法承继关系等方面对其进行评述与反思。

首先，从立法思维上看，欧盟人工智能立法体现了一种“规制-工具主义”（regulatory-instrumentalism）的立法思路，是欧盟为实现“数字欧洲议程”和“数字单一市场”战略的法律化产物。“规制-工具主义”由罗杰·布朗斯沃德（Roger Brownsword）教授提出，他将立法路径区分为“融贯主义”立法和“规制-工具主义”立法：前者属于一种传统立法思路，致力于实现法律教义的体系性与融贯性，并通过法律解释和漏洞填补的方式不断提高体系内的一致性；后者则是一种科技立法思路，指将法律规则视为实施国家政策或目标的手段与工具，法律的充分性和有效性通过是否达成相应目标来评估。[①] 相较于“融贯主义”立法而言，“规制-工具主义”更加注重通过立法实现特定规制目标，并可能在一定程度上牺牲法律清晰、简明、体系化的特点。以欧盟《人工智能法》为例，其在建设“数字单一市场”的政策背景下出台，旨在实现加强监管协同、预防市场碎片化、促进科技创新、实现技术主权和保障基本权利等多重规制目标，目标间的张力使得该法逻辑复杂、体例繁复，最终版多达400多页，极大地增加了涉人工智能企业的学习成本，提高了公众的理解门槛。

其次，从立法模式上看，欧盟立法者采取“中心化”治理方法应对人工智能风险与挑战，具体体现为一种集中式立法模式。学界对于应当采取何种方法治理人工智能存在长期争论，大致可以划分为“中心化”治理和“去中心化”治理两类观点，相对应的规制结构分别是“集中型规制”和“碎片型规制”，在立法中则分别体现为“集中式立法”和“分散式立法”。[②] 在中心化、集中型的规制结构下，人工智能具有一套完整的法律规则体系，由单一机构实施并执行法律，且往往存在一部全面性、系统性的横向立法，统一适用于各类人工智能应用领域。这种立法形式的优点在于可以避免法律间的冲突或重叠，消除分散式立法导致的“挑选法院”[③]（forum

① Roger Brownsword, *Law, Technology and Society: Re-Imagining the Regulatory Environment*, London: Routledge, Taylor & Francis Group, 2019, p. 195.

② Peter Cihon/Matthijs M. Maas/Luke Kemp, “Should Artificial Intelligence Governance Be Centralised? Design Lessons from History”, in the proceeding of AAAI/ACM AIES 2020, http://arxiv.org/abs/2001.03573，访问日期：2024-04-01。

③ “挑选法院”（forum shopping）是英美法上的术语，指当事人利用管辖权的积极冲突，在众多具有管辖权的法域或者法院中挑选对自己最有利的特定法院或者司法辖区。See *Black's Law Dictionary*, https://blacks_law.en-academic.com/11383/forum_shopping, 访问日期：2024-04-01。

shopping)现象。并且,正式和程序化的民主立法过程可以更加有效地促进政策协调与政治参与。然而,需要兼顾各方利益的集中型立法也有可能因立法妥协和让步产生"立法弱化"效应,立法进程缓慢且缺乏灵活性,若立法失败则将较难对其进行修正。[①] 相反,去中心化、碎片化的规制结构则不存在单一规制机构或者一套核心法律规则,而是具有多重规制主体,人工智能相关规范也分散在各不同层级、不同部门、不同领域的立法中,是一种纵向型、碎片化的立法形式。分散式立法有其独特优势,由于不需要制定通用型规则,其立法程序更加简便、灵活和迅捷,可以应对快速变化的社会现实,实现敏捷治理,也能够降低因立法失败带来的风险。而该立法形式的缺点也十分明显:立法的分散与碎片化也将导致规制权的分散和碎片化,这就可能造成规制重叠或者竞争现象,也可能产生不同立法间相互抵触的情形。[②] 在立法形式的选择上,欧盟立法者具有较为明显的倾向性,即采取横向的、集中型立法,其背后的一个重要考量是促进欧盟一体化进程和"数字单一市场"建设,避免立法分散导致的监管碎片化。

再次,从规制路径上看,"基于风险"的规制路径虽然旨在通过可伸缩和合比例的方法,实现人工智能时代下公民权利保障和社会风险防控的平衡,但是也存在滑入两类极端的危险。

一是"基于风险"的规制路径过于关注人工智能风险,忽视了人工智能给生产力发展和社会进步带来的收益,由此可能形成对人工智能系统的重监管模式,从而抑制创新。欧盟《人工智能法》的出台引发了产业界不少担忧,后者担心企业负担过高的监管成本,而且部分合规要求在技术上可能难以实现。一方面,以事前监管为核心的风险规制措施无疑将抬升企业合规成本。根据比利时智库欧洲国际政治经济中心(European Centre for International Political Economy)的分析预测,对科技公司进行事先监管可能导致850亿欧元的GDP损失和1000亿欧元的消费者福利损失。[③] 此外,过于严格的合规要求还将抑制小微企业、初创企业的发展,间接强化科技市场的垄断格局,因为相对于大型科技巨头而言,小微企业合规能力更弱,更需要在相对宽松的监管环境中发展成长。[④] 另一方面,立法的合规要求也可能面临技术限制。面对科技法律的兴起,布朗斯沃德提出了"法律3.0"概念,指出

① Dane Chapman, "The Ideal Approach to Artificial Intelligence Legislation: A Combination of the United States and European Union", p. 265.

② 同上, p. 289。

③ 曾雄、梁正、张辉:《欧盟人工智能的规制路径及其对我国的启示——以〈人工智能法案〉为分析对象》,第68页.

④ 考虑到小微企业、初创企业的特殊情况,欧盟《人工智能法》允许它们采取简化的方式履行某些义务,例如第9条要求的"风险管理体系"和第11条要求的技术文件。参见《人工智能法》序言146,第11条。

在法律3.0时代，法律不仅要考虑规则和政策，还需要关注技术本身，立法时既要考虑将技术手段作为法律解决方案，也要避免违背技术规律而仅出于能动性考虑制定法律规则。[①] 欧盟《人工智能法》规定了包括透明性、准确性和稳健性在内的合规规则，但在技术层面，人工智能系统难以避免地存在不透明性和偏误问题。实际上，即使是人类决策，也大量存在认知偏见和不稳定性等问题。研究证明，相比算法，人脑决策可能更加复杂和不透明。[②] 虽然欧盟尽力通过监管沙盒的方式平衡监管与创新之间的内在张力，但是从该法整体条文比例上看，该法仍然构成了以监管为主的立法格局。

二是过于强调人工智能企业和系统提供者的合规责任，使得公民、消费者和人工智能用户在风险管理中的角色被边缘化。《人工智能法》基于对风险程度、风险应对能力的整体考量，将大部分的合规责任施加给高风险人工智能系统提供者的规定具有合理性，但同时也使得人工智能终端用户和消费者的角色被边缘化。其中，终端用户和消费者几乎没有在风险管理过程中承担任何义务，且除了投诉权和人工智能自动化决策解释权，也不具有其他直接进行风险防范的权利。[③] 但事实上，人工智能系统的用户和消费者在风险产生和预防过程中占据重要地位。其一，人工智能的风险类型及程度与使用场景、运行环境密切相关。例如，在通用人工智能、生成式人工智能系统中，风险并非由模型提供者或者系统供应商预先定义，而是更多地由专业或终端用户决定。因此，有批评意见指出，欧盟《人工智能法》中基于风险水平的分级方式，并未充分考虑到人工智能风险会依据具体的运行环境而改变，从而给系统开发者带来了更多法律不确定性。[④] 其二，“供应商-用户”的相互关系和互动结构对风险合规过程也至关重要。用户可能依赖与供应商合作以遵守其法律义务，供应商也需要依靠用户的使用经验进一步优化并负责任地使用系统。[⑤]

---

① ［英］罗杰·布朗斯沃德：《法律3.0：规则、规制和技术》，毛海栋译，北京：北京大学出版社，2023年版，第1－10页。

② Cary Coglianese, “A Framework for Governmental Use of Machine Learning”, report for the administrative conference of the United States, Dec 8, 2020, p. 6, https://www.acus.gov/sites/default/files/documents/Coglianese%20ACUS%20Final%20Report%20w%20Cover%20Page.pdf, 访问日期：2024－04－01。

③ Alessandro Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, pp. 142－144.

④ N. A. Smuha/E. Ahmed Rengers/A. Harkens et al., “How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission’s Proposal for an Artificial Intelligence Act”, *working paper of SSRN Electronic Journal*, 2021－08－05, https://strathprints.strath.ac.uk/85567/1/Smuha_etal_SSRN_2021_How_the_EU_can_achieve_legally_trustworthy_AI.pdf, 访问日期：2024－04－01。

⑤ Natali Helberger/Nicholas Diakopoulos, “ChatGPT and the AI Act”, pp. 5－6.

最后，从整体构造和立法承继关系上看，欧盟《人工智能法》既延续了GDPR中“基于风险”的规制路径，又继承了《产品责任指令》(Product Liability Directive)中的产品安全监管框架，是一种双架构模式。[①] 具体表现为：其一，在立法目标层面，《人工智能法》在立法目标中对健康与安全的强调与欧盟产品安全法立法目标非常接近。其二，在机构设置层面，《人工智能法》规定由成员国国内的市场监督管理机关具体实施本法监管举措，这也与产品安全法的执法机构类似。其三，为了匹配《人工智能法》的实施，欧盟同步修订了《产品责任指令》及《机械指令》(Machinery Directive)等产品责任规则，并将《产品责任指令》的适用范围扩展到人工智能系统或者含有人工智能的商品中。[②] 前述种种均印证了欧盟人工智能立法中的“产品法”色彩。欧盟人工智能立法采取产品安全的监管框架具有其现实性和合理性。一方面，欧盟制定《人工智能法》的立法权基础来自于《欧盟运行条约》(The Treaty on the Functioning of the European Union)第114条，后者旨在建设和运行欧洲统一市场。因此，《人工智能法》被框定为一种市场调节工具，以免违反其宪法依据。另一方面，欧盟《产品责任指令》作为欧洲市场内唯一统一的产品责任法律，已经生效数十年，积累了丰富的立法经验和学术成果，这也使得欧盟人工智能立法得以从中借鉴，而非无中生有地新设规范。[③] 然而，将人工智能设限为一种“产品”并对其进行规制也存在一定局限性，因为这弱化了人工智能技术广泛存在的嵌入性特征。此外，仅从产品层面进行法律规制可能使未来很多人工智能脱离法律监管框架。

(二)对我国人工智能立法的启示与借鉴

他山之石，可以攻玉。在我国，制定“人工智能法”已成为未来趋势，但如何立法还存在很大争议。欧盟《人工智能法》作为世界人工智能立法领域的先驱者，其立法经验和成果无疑对我国未来人工智能法的制定具有启示与借鉴意义。

在立法功能上，推动人工智能立法有助于整合当前碎片化、分散化和区域性的人工智能治理格局，参与全球治理，提升我国法治软实力。从欧盟经验来看，人工智能立法所负载的一个重要功能预设在于市场整合与监管协同功能，即整合区域性、碎片化的监管市场，促进欧盟内商品服务的在线流通。这对我国人工

① Almada Marco/Petit Nicolas, “The EU AI Act: A Medley of Product Safety and Fundamental Rights”, Working Paper of European University Institute, pp. 7 - 27, here p. 12, https://hdl.handle.net/1814/75982, 访问日期:2024 - 04 - 01。

② Novelli and others, “Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity”, *working paper of SSRN Electronic Journal*, 2024 - 02 - 19, https://arxiv.org/abs/2401.07348, 访问日期:2024 - 04 - 01。

③ 同注①，p. 26。

智能治理也至关重要。现阶段，我国人工智能治理格局呈现出碎片化、分散化和区域性特征，这主要体现在三个方面。第一，人工智能治理缺乏整体布局，规范层级低、效力弱。在涉及人工智能的规范中，地方性立法先于全国性立法、非正式法律渊源多于正式法律渊源，这会导致监管区域化和碎片化。① 第二，人工智能治理注意力分配不均，规范分布呈现拼盘式、补丁式特征。人工智能由数据、算法、算力三项智能要素构成，在我国涉及智能要素的规范格局中，数据类规范数量多，算法、算力类规范数量少，政策制定者和立法者对后两项要素的关注不足。此外，对人工智能的单独规范呈补丁式特征，如生成式人工智能等受关注的类型存在单独规范，而其他人工智能类型则缺少布局。第三，人工智能治理体系扁平化，基础性技术标准、道德伦理标准供给严重不足。标准、规则和指令支撑了人工智能治理体系，构筑出复合立体的治理大厦。其中，标准具备敏捷性、技术性和可通约性，能够有效引导人工智能技术有序、协同、稳健发展，但目前国家层面的人工智能技术和伦理标准十分缺乏、立法进展缓慢、技术标准化程度不足等问题，使得治理体系失衡。② 因此，人工智能立法是我国人工智能治理实践所需和未来趋势。一方面，人工智能立法能够填补人工智能治理在法律层级的规范缺失，促进监管协同，避免出现监管地方化和市场碎片化；另一方面，人工智能立法也能明确我国人工智能发展方向和基本原则，有助于形成均衡发展、复合立体、分层分级的高效治理体系。

此外，我国在立法过程中还应当避免将人工智能立法与促进科技创新对立。目前产业界对立法的疑虑在于担忧立法将抬升企业成本、构筑监管高地、造成过度监管并抑制科技创新，其背后逻辑在于将创新与监管对立起来。将立法规制和产业发展相对立的观点并不罕见，例如，有研究者认为，相比于欧洲、日本和韩国，美国互联网平台发展更为繁荣的原因在于其更为宽松的法律规定和更加具有“平台友好”特点的政策环境。③ 这类认识有其合理之处，但并不全面。例如，周汉华指出，在个人信息保护领域，我国可以探索一条保障与发展激励相容的治理之道；④ 曾雄等人也认为，政府立法和规制行动对产业发展的实际影响不能跳过扎实的实

① 陈亮：《人工智能立法体系化的困境与出路》，载《数字法治》，2023年第6期，第10－17页，这里第13－14页。

② 张欣：《我国人工智能技术标准的治理效能、路径反思与因应之道》，载《中国法律评论》，2021年第5期，第79－93页，这里第84－85页。

③ 参见曹建峰：《论互联网创新与监管之关系——基于美欧日韩对比的视角》，载《信息安全与通信保密》，2017年第8期，第72－80页，这里第79页。

④ 周汉华：《探索激励相容的个人数据治理之道——中国个人信息保护法的立法方向》，载《法学研究》，2018年第2期，第3－23页，这里第3页。

证研究而直接下定论。①

在立法路径上，我国未来人工智能立法适宜采取“基于风险”与“基于权利”相结合的规制路径，并以原则治理为补充，构建起层次分明、体系合理的人工智能法律体系。“基于风险”与“基于权利”两条规制路径并非截然对立、非此即彼的关系，而是相辅相成、互为补充，共同作用于人工智能的全生命周期治理。其一，“基于风险”和“基于权利”的规制路径的侧重点与实施阶段不同。前者着重事前预防和事中阻击，旨在提供一种系统性的治理方案；后者本质上是通过赋权实现事后控制，以公民行权的方式达到一种私人执法的效果并提供救济。但是，无论单独适用哪一种立法路径都存在局限性。例如，风险路径的缺陷在于对个体性权利的关照不足，事后救济渠道阻滞；而权利路径的不足则在于难以改变权利与权力间的势差，个体怠于行权将导致对违法行为约束乏力，并且在事前和事中阶段无法实现有效治理。其二，权利与风险本身属于不同的知识和社会组织领域。权利概念来源于法律实践，基于事先确定的法定权利，公民可以在违法事实发生后，通过向法院提起诉讼请求的方式行权，间接达到控制违法的效果；风险则常与风险管理实践相联系，后者则更多地出现在公司治理语境下，能够通过客观、中立、科学的风险评估的方式来降低不利后果产生的概率。② “风险评估的主要任务是探索并识别与事件相关的负面后果的类型、强度和可能性。”③因此，风险评估和管理的基本过程从根本上并不涉及权利的性质，而是涉及某些后果发生的可能性。风险逻辑与权利逻辑的交汇出现在新兴科技立法领域，一方面，立法吸纳了统计学和科学上的风险概念；另一方面，“针对特定不利后果的风险”也转变为一种“针对权利的风险”。综上，风险路径与权利路径虽然存在异质性，但并非不可兼容，相反，唯有将二者结合方能造就完整的人工智能法治体系。

实际上，在目前已出台的学者建议稿和人工智能倡议中，我们已经可以看到风险路径与权利路径并存的特征。例如，《中华人民共和国人工智能法（学者建议稿）》以专章的方式规定了人工智能使用者的平等权、知情权、隐私与个人信息保护权、人工智能决策解释与拒绝权等，并规定公民可以通过提起诉讼的方式行权。④又如，《人工智能示范法2.0（专家建议稿）》规定了公民针对人工智能民事侵权行

① 曾雄、梁正、张辉：《欧盟人工智能的规制路径及其对我国的启示——以〈人工智能法案〉为分析对象》，第71页。

② See Niels Van Dijk/Raphaël Gellert/Kjetil Rommetveit, “A Risk to a Right? Beyond Data Protection Risk Assessments”, *Computer Law & Security Review*, Vol. 32, No. 2, 2016, pp. 286 – 306, here pp. 290 – 291.

③ Ortwin Renn, *Risk Governance Coping with Uncertainty in a Complex World*, London: Sterling, 2008, p. 5.

④ 参见《中华人民共和国人工智能法（学者建议稿）》第三章，《使用者权益保护》，https://sdbdra.cn/newsinfo/6966473.html，访问日期：2024 – 04 – 23。

为的诉权(第 70 条)，以及针对主管机关的监管行为提起复议或诉讼、督促国家机关履行保护职责的权利(第 72 条)。[①] 除了赋权条款，前述专家建议稿和我国 2023 年 10 月发布的《全球人工智能治理倡议》都不约而同地纳入了风险分级分类、风险管理评估等制度。例如，《全球人工智能治理倡议》提出要推动建立风险等级测试评估体系，实现敏捷治理及分级分类管理；《人工智能示范法 2.0(专家建议稿)》规定人工智能提供者应当完成风险识别、评估和管理义务(第 41 条)等。

应当注意的是，兼采风险路径和权利路径并非二者的简单相加，而是要使二者有机结合、相互配合。具体而言，首先，我国未来人工智能立法要明确赋予公民何种权利，该权利作用于人工智能治理的哪一环节，应通过民事诉讼还是行政执法的方式对其进行保障。其次，我国应明确采取何种风险分级分类标准与方式，例如《中华人民共和国人工智能法(学者建议稿)》将人工智能区分为一般人工智能和关键人工智能，《人工智能示范法 2.0(专家建议稿)》则采取负面清单的方式予以区别。再次，我国应建立行之有效的权利影响评估制度。科技立法中的风险本质上是一种针对权利的风险，而非仅仅针对不利后果的风险。例如，欧盟《人工智能法》除了规定风险管理体系，还建立了基本权利影响评估制度(第 27 条)，我国《个人信息保护法》也规定了个人信息影响评估制度(第 56 条)，其目的在于保障公民基本权利、个人隐私免受自动化智能决策和数据处理行为的不良影响。我国未来人工智能立法也可以建立类似的权利影响评估机制，通过相对明确的技术框架和指标体系，保护公民权利免受具有隐蔽性、不可察觉的人工智能自动化决策的侵扰。最后，我国人工智能立法还要避免欧盟人工智能立法中存在的一些“短板”。例如，欧盟试图制定一份全领域人工智能系统的集中式横向型立法，这不可避免地导致了立法进程缓慢；此外，对于多方妥协的追求还可能导致立法弱化问题。在数字时代，人工智能等高新科技治理不仅需要法律的明确性，也需要法律的适应性和敏捷性，过于厚重的立法更易产生滞后性，从而导致立法失败。因此，我国更好的立法进路是先制定框架性、原则性、总括性的人工智能基本法，然后再授权各地方、部门制定单行法规范，从而形成复合立体、层级分明的人工智能规范体系。又如，欧盟《人工智能法》在对人工智能风险进行分级和分类时，以系统提供者预设的使用目的和应用场景为标准，忽视了系统用户和消费者的作用与影响，将用户和消费者的角色边缘化。因此，我国在未来人工智能立法时应当避免基于预先设定类型进行风险划分的静态式风险观，应结合人工智能使用目的、使用场景进行动态考量，并关注人工智能使用中“用户-供应商”关系，打造多元主体共建、公私法兼具的人工

① 参见《人工智能示范法 2.0(专家建议稿)》，2024-04-22，https://aisg.tongji.edu.cn/info/1005/1211.htm，访问日期：2024-04-25。

智能治理体系。

结 语

欧盟《人工智能法》填补了世界人工智能治理体系中法律规范的缺失，具有重大意义，是一次有益的立法尝试，同时也深刻影响着全球人工智能治理和监管格局。在我国，制定一部专门的人工智能法也成为必然趋势，但对于如何立法尚未形成统一认识。他山之石，可以攻玉。通过还原欧盟人工智能立法的演进历史和规制路径，我们可以发现其经历了从政策到软法再到硬法的渐次发展过程，立法中蕴含了“基于原则”“基于权利”和“基于风险”三种相互区别又互为补充的规制路径。该法的风险分级分类、风险管理系统、合格评价、沙盒监管制度设计，为我国日后立法提供了可供借鉴的经验与资源。人工智能立法牵涉甚广，并不存在放之四海而皆准的定律，我们应当辩证地看待欧盟立法中的成败得失，结合我国人工智能产业发展情况和已有监管资源，制定出科学、合理、适应技术特征和时代发展需求的人工智能立法方案。通过对欧盟立法的借鉴与反思，本文提出我国未来人工智能立法应当采取“基于权利”与“基于风险”相结合的法律规制路径，并以原则治理为补充，构建起层次分明、体系合理的人工智能法律体系。

责任编辑：郑春荣