人工智能监管的路径选择

——欧盟《人工智能法》的范式、争议及影响*

王天凡

内容提要:中国的人工智能立法已提上日程,而该立法应遵循何种监管路径尚无定论,不同法域的前期探索均值得参考。欧盟《人工智能法》是欧洲数字立法版图展开的重要一步,是欧盟人工智能战略的基础性法律。欧盟《人工智能法》采取横向立法模式,选择了"基于风险"的监管方法,将人工智能系统分为四个不同的风险级别,对其设置相应强度的监管措施:被禁止的人工智能系统、受详细规则约束的高风险人工智能系统、其他人工智能系统,以及通用人工智能模型。其中,公共场所"实时"远程生物特征识别及通用人工智能模型监管等问题备受争议。欧盟《人工智能法》面临标准制定、市场主体成本增加、实施成本高昂等问题。欧盟与美国、英国等代表性国家对人工智能的规制路径各不相同。中国应立足于本国人工智能战略,选择横纵结合的监管路径。

关键词:欧盟　《人工智能法》　通用人工智能　生物特征识别　监管沙盒

人工智能及其监管对人类而言是新生事物,世界各主要法域都在尝试制定监管规则以应对这一挑战。中国《国务院 2024 年度立法工作计划》也已将人工智能立法列入预备提请审议项目。目前,立法应采用何种监管路径尚无定论,而各法域的前期探索对这一问题具有参考价值。欧盟《人工智能法》(Artificial Intelligence Act, AI Act)的制定集合了欧盟范围内科技界、产业界、宗教界、法学界、哲学界、政治界等诸多领域及人群的代表性观点、权衡以及妥协,也在很大程度上反映了欧盟民众对待人工智能的社会心理,可以说是最前沿、最具代表性的立法尝试之一。深入研究欧盟人工智能立法的背景、监管范式、规制要点及立法过程中的核心争议等,有助于学界理解人工智能法律规制的核心问题,对中国人工智能的立法和研究亦有所裨益。

* 本文为国家社会科学基金"人工智能时代合同法制度的体系性更新研究"(项目编号:23BFX180)的阶段性研究成果。

一 欧盟数字立法与《人工智能法》

(一)欧盟的数字立法版图

从历史图景来看,欧洲私法为了适应社会数字化发展经历了三波浪潮:①第一波始于 20 世纪 90 年代末,旨在使法律规则适应电子商务的要求;第二波目前方兴未艾,主要是为了应对新技术条件下商业模式的法律规制,包括对数字内容或服务和智能产品交易的规制;第三波则初露端倪,主要聚焦于私法如何应对自主运行的人工智能。②

在战略层面,欧盟将 2020—2030 年定为“数字十年”,③制定了明确的数字战略目标,指出了提高数字技能、实现行政和商业数字化、促进研究和创新、缩小数字差距以及升级数字基础设施等发展方向。与此同时,欧盟在个人信息保护、数字市场治理和数据资源利用等数字化发展的关键领域不断推进立法,意图在数字化未来的法律规制方面发挥领导作用。

在个人信息保护、数据隐私和数字身份方面,自 2018 年《一般数据保护条例》(GDPR)生效后,欧盟委员会于 2021 年提出《数字身份条例框架》草案(eID),据此,所有欧盟公民、居民和企业都可以使用欧洲数字身份钱包。新框架修订了 2014 年生效的欧盟《内部市场电子交易的电子识别和信任服务条例》(eIDAS),这一条例为欧盟安全获取公共服务和进行在线跨境交易奠定了基础。2023 年,欧盟理事会和欧洲议会就欧盟数字身份问题达成临时协议,这标志着欧盟朝着“2030 数字十年”公共服务数字化目标迈出了重要一步,被认为是欧盟成为数字领域全球参考样本、保护欧盟的民众权利和价值观的关键进展。④

在数字市场治理方面,随着 2022 年《数字市场法》和《数字服务法》的通过,欧盟在数字平台监管与数字市场竞争方面树立了监管标杆。前者对大型平台课以“守门人”义务,防止其对其他企业和消费者施加不公平条件。后者对施行逾 20 年的欧盟《电子商务指令》(2000/31/EC)进行了重大修订,从内容管理、广告推送、透明度等方

① Dirk Staudenmayer, HdB Europäisches Digitales Zivilrecht, 2023, §§ 2 und 25.

② Dirk Staudenmayer, Haftung für Künstliche Intelligenz, Die deliktsrechtliche Anpassung des europäischen Privatrechts an die Digitalisierung, NJW 2023, 894.

③ Commission Work Programme 2024, “Delivering Today and Preparing for Tomorrow,” COM(2023) 638 final, Strasbourg, 17.10.2023.

④ Nadia Calviño, “In European Digital Identity: Council and Parliament Reach a Provisional Agreement on eID,” https://www.consilium.europa.eu/en/press/press-releases/2023/11/08/european-digital-identity-council-and-parliament-reach-a-provisional-agreement-on-eid/.

面构建数字平台“阶梯式”监管模式,力图使欧盟的单一市场在数字领域更加体现公平、消费者友好和竞争力。

在数据资源的利用方面,2022 年,欧盟委员会提出了欧盟《数据法案》(Data Act)草案,并于 2023 年 12 月正式获得通过并发布。该法案是继 2022 年的《数据治理法案》(Data Governance Act)后,欧盟委员会在“欧洲数据战略”①之下的第二项主要立法举措。欧盟《数据法案》规定了数据持有者在符合条件时向用户和第三方提供数据的义务、向公共部门和机构提供数据的义务、用户在不同的数据处理提供商之间切换的权利、非个人数据的国际传输、禁止不公平合同条款等。该法案旨在通过促进跨部门和利益相关者的数据共享,确保数字环境中参与者之间数据价值分配的公平性,刺激数据市场竞争,使欧盟成为数据驱动型社会的领导者。

(二)《人工智能法》的立法背景:欧盟的人工智能战略

作为已经展开的数字立法版图的“第三步”,欧盟在过去五年间密集出台了数十份涉及人工智能的法规(Regulation)、指令(Directive)、准则(Guideline)、决议(Resolution)等。尤其在 2020 年之后,相关内容更是呈现“井喷”之势,②尽显欧盟对人工智能关注之密切。欧盟对人工智能技术的态度在此期间也发生了转变,由一边倒的“促进”和“推动”转变至“促进与规制并举”。

2018 年,欧盟委员会先后出台了两份关于人工智能的重磅文件——“欧洲人工智能战略”及“人工智能协调计划”。二者的核心目标均在于促进人工智能在欧洲的发展。在“欧洲人工智能战略”③中,欧盟委员会指出,人工智能是欧盟委员会“工业数字化战略”④和“新产业政策战略”⑤的一部分,计算能力的增长、数据的可用性和算法的进步使人工智能成为 21 世纪最具战略意义的技术之一。欧洲领导人已将人工智能列为首要议程,认为欧盟应该采取协调一致的方法,充分利用人工智能提供的机遇并应

① Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, “2030 Digital Compass: The European Way for the Digital Decade,” COM(2021) 118 final, Brussels, 9.3.2021, https://eufordigital.eu/wp-content/uploads/2021/03/2030-Digital-Compass-the-European-way-for-the-Digital-Decade.pdf.

② 限于篇幅,本文无法全面列举细述欧盟五年间所有针对人工智能的文件,以下仅为举要。

③ Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, “Artificial Intelligence for Europe,” COM(2018) 237 final, Brussels, 25.4.2018.

④ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, “Digitising European Industry Reaping the Full Benefits of a Digital Single Market,” COM(2016) 180 final, Brussels, 19.4.2016.

⑤ Communication from the Commission, “Investing in a Smart, Innovative and Sustainable Industry, a Renewed EU Industrial Policy Strategy,” COM(2017) 479 Final, Brussels, 13.9.2017.

对其带来的新挑战。欧洲保持竞争力的主要挑战之一是确保人工智能技术在其工业经济中的应用。欧盟委员会认为,欧洲可以凭借其价值观和工业优势,包括利用世界一流的研究人员、实验室和初创公司、数字单一市场、丰富的工业、研究和公共部门数据等,在开发和使用人工智能方面发挥带头作用。“欧洲人工智能战略”的三大支柱分别是:增加对人工智能的公共和私人投资,为社会经济变革做好准备,以及确保人工智能的开发和应用在符合欧洲价值观的框架内进行,包括尊重基本权利、伦理规则、透明度及责任承担等。这一战略制定了雄心勃勃的目标:欧盟必须加大对人工智能的公共和私人投资,以实现未来十年每年投资超过 200 亿欧元的目标。①

欧盟委员会在“人工智能协调计划”②中指出,欧洲目前在人工智能领域的私人投资方面处于落后地位。如果不做出重大努力,欧盟就有可能失去人工智能带来的机遇,面临人才流失的挑战,只能成为其他国家开发的解决方案的消费者。而监管机构的重要职责则在于消除分散市场造成的障碍,避免人工智能等战略领域的市场碎片化。该计划在政策层面上包括:建立欧洲人工智能公私合作模式,为初创企业和创新型中小企业提供更多融资;提高国家研究能力,促进欧洲最好的研究团队之间的合作;促进人工智能在经济中最广泛的应用;调整教育与职业培训体系,吸引人才,支持劳动力市场转型;以《一般数据保护条例》为支柱,建立包括公共部门在内的对人工智能发展至关重要的欧洲数据空间,解决数据访问、非个人数据流动等问题,并以“欧洲高性能计算计划”(EuroHPC)③和“欧洲处理器计划”④为保障;制定具有全球视野的伦理准则,确保相关法律框架有利于创新、人工智能应用和基础设施的安全,并促进国际安全。

---

① 在该战略制定的前一年(2017 年),欧盟在人工智能方面的公共和私人研发投资已达到 40 亿至 50 亿欧元。该战略声称,整个欧盟(公共和私营部门合计)应致力于到 2020 年年底将这一投资增加到至少 200 亿欧元,并保持在未来十年内每年均超过这一数额。欧盟委员会建议,在 2021—2027 年的下一个计划编制期间,欧盟每年从“地平线欧洲”(Horizon Europe)和“数字欧洲”(Digital Europe)计划中为人工智能投资至少 10 亿欧元。“地平线欧洲”计划被认为是有史以来最雄心勃勃的欧盟研究和创新框架计划,旨在支持欧洲人工智能战略。关于欧洲数字计划的财务限额,参见 Regulation (EU) 2021/694 of the European Parliament and of the Council of 29 April 2021 Establishing the Digital Europe Programme and Repealing Decision (EU) 2015/2240 (Text with EEA Relevance)。

② Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, “Coordinated Plan on Artificial Intelligence,” COM(2018) 795 final, Brussels, 7.12.2018.

③ “European High-Performance Computing Initiative,” https://ec.europa.eu/digital-single-market/en/eurohpc-joint-undertaking. 2024 年 1 月 24 日,欧盟委员会公布了一项关于“修订(欧盟)第 2021/1173 号条例的理事会条例提案,涉及为初创企业提供 EuroHPC 计划,以提高欧洲在可信人工智能领域的领导地位”的提案,可见欧盟在此领域的进展。

④ “European Processor Initiative,” https://eurohpc-ju.europa.eu/research-innovation/our-projects/european-processor-initiative-epi_en.

2019 年由欧盟人工智能高级别专家组起草的“值得信赖的人工智能伦理准则”,[①]正是为回应前述文件中对人工智能监管之需要而拟定。这标志着欧盟对人工智能的规制开始“加码”。该准则提出,要实现“可信赖的人工智能”,必须具备三个要素:(1)合法;(2)符合道德原则;(3)稳健。以此为核心,该准则列明了七项关键要求:人工机构和监督、技术鲁棒性和安全性、隐私和数据管理、透明度、多样性和非歧视及公平性、环境和社会福祉、问责制。需要注意的是,该准则不具有约束力,因此不会产生新的法律义务。但欧盟的许多现有法律已经反映了其中一项或多项关键要求,例如安全、个人数据保护、隐私或环境保护等。

2020 年,欧盟委员会发布《人工智能白皮书》。[②] 白皮书中明确指出,尽管在发布上述伦理准则时采用了软法方式,但欧盟委员会认为,适用于人工智能系统的道德准则和现有法律已不足以解决人工智能开发和部署所带来的风险,因此,欧盟决定转向强制立法,并呼吁制定一项新的针对人工智能的强制性法规。白皮书重申“人工智能应该为人类服务,成为社会造福的力量”,其指导原则是创建卓越生态系统(“政策框架”)和信任生态系统(“监管框架”)。在规制的强度方面,白皮书认为应以“有效性”和“不过分规范”为标准。虽未涉及具体措施,但白皮书认为应遵循基于风险的方法,确保干预适当并足够灵活,以适应技术进步并精确提供法律之确定性。但该白皮书也遭到了诸多批评,如有观点认为白皮书以及整个欧盟的战略模棱两可,缺乏远见。[③]

2022 年,欧盟委员会发布了《人工智能责任指令》草案,[④]建议引入专门针对人工智能系统所造成损害的新规则,对欧盟责任框架进行补充和现代化,包括构建“同等保护”原则、“因果关系推定”规则等。这一指令出台的直接原因是,鉴于人工智能侵权案件与传统侵权案件的不同,欧盟各国国内法院在法律适用上面临困境和不统一,如果欧盟不及时采取行动,成员国必将各自调整其国家责任规则,这会进一步加剧该领域法律的碎片化,损害欧盟单一市场,并导致跨境贸易企业和中小企业成本过高。另外,鼓励对新兴数字技术的信任,并为欧盟人工智能产品和服务的发展创造必要的

① European Commission, “Ethics Guidelines for Trustworthy AI,” https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

② European Commission, “White Paper on Artificial Intelligence—A European Approach to Excellence and Trust,” COM(2020) 65 final, Brussels, 19.2.2020.

③ Emre Kazim and Adriano Koshiyama, *Lack of Vision: A Comment on the EU's White Paper on Artificial Intelligence*, SSRN 2020, https://ssrn.com/abstract=3558279.

④ “Proposal for a Directive of the European Parliament and of the Council on Adapting Non-contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive),” COM(2022) 496 final, Brussels, 28.9.2022.

投资稳定性,也是欧盟委员会提议背后的重要动因。①

上述针对人工智能系统相关责任问题的新规则只是欧盟一系列举措中的一部分,而通过《人工智能法》制定适用于所有在欧盟市场投放或使用的人工智能系统的共同规则,正是欧盟关于人工智能法律规制的重点。

(三)《人工智能法》的立法进程

早在 2021 年 4 月,欧盟委员会就提出了全球首个人工智能法律框架提案②——欧盟《人工智能法》草案。为使“欧洲成为值得信赖的人工智能全球中心”,③该草案充分回应了上述欧洲议会及理事会的多项呼吁,包括解决某些人工智能系统的不透明性、复杂性、偏差、一定程度的不可预测性和部分自主行为等问题。④

2024 年 1 月,欧洲议会正式发布了《人工智能法》草案的修正案(以下称为“一读修正案”)。⑤ 一读修正案中包含了大量修改,这是三方会谈密集谈判的结果,也是对欧盟委员会草案所受到的各种批评的回应。⑥ 2024 年 3 月,欧洲议会正式通过了《人工智能法》。议会通过版在一读修正案基础上又做出大量修改,最终于 2024 年 5 月得到欧盟理事会的批准。⑦

作为一部监管人工智能的全面性、强制性法律,《人工智能法》的立法目标在于:(1)确保进入欧盟市场和使用的人工智能系统是安全的,并尊重有关基本权利和符合欧盟价值观的现行法律;(2)实现法律的确定性,以促进人工智能领域的投资和创新;(3)加强对适用于人工智能系统的基本权利和安全要求的现行法律的管理和有效执

---

① 欧盟委员会 2020 年公布的一项关于人工智能技术使用情况的调查得出的结论是,33%的企业认为,潜在损害的责任是欧盟采用人工智能的主要外部挑战之一。European Commission, Directorate-General for Communications Networks, Content and Technology, “European Enterprise Survey on the Use of Technologies Based on Artificial Intelligence,” Publications Office of the European Union, 2020.

② David Bomhard und Marieke Merkle, Europäische KI-Verordnung, Der aktuelle Kommissionsentwurf und praktische Auswirkungen, RDi 2021, 276.

③ Europäische Kommission, Für vertrauenswürdige Künstliche Intelligenz: EU-Kommission legt weltweit ersten Rechtsrahmen vor, https://germany.representation.ec.europa.eu/news/fur-vertrauenswurdige-kunstliche-intelligenz-eu-kommission-legt-weltweit-ersten-rechtsrahmen-vor-2021-04-21_de.

④ Council of the European Union, Presidency Conclusions, “The Charter of Fundamental Rights in the Context of Artificial Intelligence and Digital Change,” 11481/20, 2020.

⑤ Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206-C9-0146/2021- 2021/0106(COD))。为示区分,本文将 2021 年由欧盟委员会发布的草案称为“原草案”或“委员会版草案”。

⑥ Daniel Feuerstack, Daniel Becker und Nora Hertz, Die Entwürfe des EU-Parlaments und der EU-Kommission für eine KI-Verordnung im Vergleich, ZfDR 2023, 421.

⑦ 欧洲议会于 2024 年 3 月通过,并最终得到欧盟理事会批准的条例在内容上基本未作修改,为与“草案”及“一读修正案”比较来说明条例内容变动节点,下文将这一版本称为“通过版”。

行;(4)促进合法、安全和可信的人工智能应用单一市场的发展,防止市场分散。在规则的设置上,草案体现出“平衡”与“适度”的监管思想,将干预限制在解决人工智能相关风险和问题的最低必要性之上,避免过度限制或阻碍技术发展,或不成比例地增加将人工智能技术投放市场的成本。①

由此,欧盟建构了以《人工智能法》为核心,包括《人工智能责任指令》《产品责任指令》等以特别问题为规制对象的指令,针对平台但也涉及人工智能的法规(如《数字服务法》和《数字市场法》等),以及不涉及具体人工智能技术的法规(如《一般数据保护条例》等)作为支柱的对人工智能系统的监管体系。

欧盟对待人工智能技术虽然打着“强规制”的旗号,其背后的根本态度依然是“促进”。其“强规制”的实际目标在于促进人工智能技术达至“可信”,从而为欧盟企业、公共部门及民众对人工智能的接纳铺平道路,拓展欧盟的人工智能技术市场及数据空间;远期目标则关系到欧盟的整体战略、核心利益、价值观和话语权的实现。

二 规制路径与要点:欧盟《人工智能法》的突破性尝试

(一)监管范式

从技术上看,一方面,《人工智能法》制定了一个强大的法律框架:它根据“基于风险的方法”构建相应的“横向”监管体系,在调整对象“广泛性”的基础上,尽力实现“确定”而“适度”的监管;另一方面,它的基本监管选择又必须是面向未来的,包括人工智能系统应遵守的原则性要求,必须尽力使规制具有“灵活性”,为技术的发展留足空间。

1.“横向监管”——规制对象的广泛性

《人工智能法》是有史以来首次尝试对人工智能进行横向监管的法规,适用于在欧盟市场上投放或使用的几乎所有的人工智能系统。②

在《人工智能法》正式形成之前,欧盟委员会曾拥有四种不同程度的监管干预的

① Proposal for a Regulation of the European Parliament and of the Council, “Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts,” COM(2021) 206 final, Explanatory Memorandum, Brussels, 21.4.2021.

② European Parliament, Artificial Intelligence Act, Briefing, https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

政策选择方案,[①]其中方案二是建议采取部门性的“临时”(“ad-hoc”)方法,即“纵向”(vertical)的监管思路,通过“定制”方式,针对不同的人工智能应用或类型分别制定法规。这种思路最终被欧盟立法者抛弃,轻而采取了横向监管的立法工具。

这两种立法模式各有优劣,欧盟的选择更多考虑的其实是作为联盟层面立法的目标和现实需求。与其他欧盟层面数字立法相同,欧盟《人工智能法》的基本诉求之一是促进合法、安全和可信的人工智能应用单一市场的发展,防止市场分散和各国分别立法导致的法律碎片化。在构建统一化规则这一点上,横向立法具有绝对的优势。同时,对于欧盟而言,实际的问题是其立法商谈成本畸高,采取行业或部门立法型的纵向立法模式并不现实。欧盟的立法者无法像一个主权国家的立法者那样,及时关注各类型(包括新出现的)人工智能应用并做出响应,制定相应的垂直法规。采取横向监管的方法,意味着在某种程度上节约了立法的成本,而将更实质和具体的工作,如标准制定等,留给了各相关标准制定者、具体监管机构和法院等。

作为欧盟《人工智能法》横向监管的核心要素,“人工智能”的定义直接决定了该法调整对象的范围。条例中界定的“人工智能系统”是指一种基于机器的、被设计为以不同程度的自主方式运行的系统,在部署后可表现出适应性,并且出于明确或隐含的目标,从其收到的输入信息中推断出如何生成输出,如可以影响物理或虚拟环境下的预测、内容、建议或决策。这一定义与原草案相比有三个最明显的修改:第一,强调人工智能系统必须具有“不同程度的自主性”,这意味着人工智能系统至少在一定程度上独立于人类控制,能够在没有人类干预的情况下运行。[②] 所谓“自主性”强调须排除仅基于自然人定义的规则而自动执行操作的系统。在原草案公布后,人工智能的定义条款遭到猛烈抨击。学者认为,这一定义十分不专业,过于宽泛,未能触及人工智能技术与其他一般技术之间的核心区别,[③]据此,智能电表和其他基于人定规则的系统等几乎所有软件都可能被包括在内。[④]修改后加入的这一限缩条件显著缩小了定义的

① Proposal for a Regulation of the European Parliament and of the Council, “Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts”.

② 这一修改在一读修正案中即已确定,并最终得以保留。参见一读修正案鉴于条款(6),Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206-C9-0146/2021—2021/0106(COD)), 以及最终版鉴于条款(12)。

③ 2021 年欧盟委员会版草案中的人工智能定义为:“人工智能系统”(AI system)是指采用附件一所列的一种或多种技术和方法开发的软件,能够针对人类定义的特定目标,生成内容、预测、建议或决定等输出,影响与其互动的环境。

④ Philipp Hacker und Amelie Berz, Der AI Act der Europäischen Union—Überblick, Kritik und Ausblick, ZRP 2023, 226.

范围。

第二,增加了“在部署后可表现出适应性”要件。人工智能系统在部署后表现出的适应性是指自我学习能力,允许系统在使用过程中发生变化。① 这是针对一读修正案中“人工智能”概念所受到最大批评的回应。学者尖锐地指出,其未能体现核心要素:“学习”及“适应新环境”的能力。否则,电动牙刷的系统也可能被归属于定义中的“自主”系统,而其实际上距离人工智能还很远;若不加入这一要素对概念进行限制,《人工智能法》将成为“软件法案”。②

第三,明确人工智能系统具有“推理”功能。条例③鉴于条款中(第 12 条)特别说明,人工智能系统的概念应该基于该系统的关键特征,以区别于更简单的传统软件系统或编程方法,而人工智能系统的一个关键特征是它们的推理能力。这种推理能力是指获取可能影响物理和虚拟环境的输出(如预测、内容、建议或决策)的过程,以及人工智能系统从输入或数据中推导出模型或算法的能力。人工智能系统的推理能力超越了基本的数据处理,可以进行学习、推理或建模。④

除此之外,条例的另一个重要修正是删除了原草案中对人工智能技术和方法进行补充界定的附件一。⑤ 这也是对批评意见的回应,即认为附件一不适当地扩大了人工智能系统的定义,使得传统的非智能软件(如 Excel 电子表格中的简单程序)也被涵盖在人工智能系统的概念范畴,因为其可以“在一定程度上”自主运行。⑥ 相对的,条例最终淡化了对人工智能技术和方法的强调,仅在鉴于条款提及,人工智能系统实现推理的技术包括从数据中学习如何实现某些目标的机器学习方法,以及从待解决任务的

① Whereas 12, European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206-C9-0146/2021-2021/0106(COD)).

② Philipp Hacker und Amelie Berz, Der AI Act der Europäischen Union—Überblick, Kritik und Ausblick, ZRP 2023, 226.

③ 以下若无特别说明,均以“本条例”或“条例”指称欧盟《人工智能法》。

④ European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206-C9-0146/2021-2021/0106(COD)).

⑤ 原草案在附件一中规定:“人工智能技术和方法”包括(a)机器学习方法,如监督学习、无监督学习和强化学习以及使用包括深度学习在内的各种方法;(b)基于逻辑和知识的方法,如知识表示、归纳(逻辑)编程、知识库、推理和演绎引擎、(符号)推理和专家系统;(c)统计方法、贝叶斯估计、搜索和优化方法。该附件在一读修正案中即被完全删除。

⑥ Daniel Feuerstack, Daniel Becker und Nora Hertz, Die Entwürfe des EU-Parlaments und der EU-Kommission für eine KI-Verordnung im Vergleich, ZfDR 2023, 421.

编码知识或符号表示中进行推理的基于逻辑和知识的方法。①

对“人工智能系统”这一核心概念的界定,直接关系到横向立法模式下欧盟《人工智能法》的调整范围和规制对象的问题。当然,基于其抽象性和概括性的表述,欧盟《人工智能法》对“人工智能系统”规定的构成要件也是相对灵活的,这在一定程度上缓解了纯粹横向框架的广泛性带来的关键精度挑战,使得合规策略能够在跨部门和技术发展的过程中保持适当的灵活性。②

2.“基于风险”的方法——确定性与灵活性

《人工智能法》最突出的特点,即其“基于风险”的监管方法(risk-based approach to regulation,RBR)对现有人工智能系统进行分类,并据此“精确”规定不同的要求和义务,使干预措施与风险水平成正比。条例在肯定人工智能系统的技术中立的基础上,依据所涉及的权利和利益的不同,区分了四个不同的风险级别,相对应的监管也分为四级,采取逐步降级的干预措施,从完全禁止到自由放任。四个风险级别具体如下:具有不可接受风险的被禁止的人工智能系统、受详细规则约束的高风险人工智能系统、仅受某些透明度要求约束的低风险人工智能系统,以及从一读修正案开始作为一个新的风险级别被引入的通用人工智能,包括生成式人工智能模型,例如 ChatGPT。

实际上,欧盟委员会在本条例制定之初尚有其他替代政策选择,除了上述提及的纵向部门性“临时”方案之外,还包括自愿标签计划及对所有人工智能系统提出强制性要求的横向方法。为何欧盟委员会最终确立“基于风险”的监管路径?如果我们稍作回顾,会发现这种规制模式在欧盟立法中并不鲜见。从早期针对食品安全、医疗设备和药品等领域的规制,③到近年来的《数字服务法》等数字领域立法,均采用了此类监管路径。④ 欧盟委员会在确立首选方案之前对每个政策选项都根据经济和社会影响进行了评估,最终确定首选方案是“方案 3+”,即仅针对高风险人工智能系统设置强

① European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206-C9-0146/2021-2021/0106(COD)).

② Matt O' Shaughnessy and Matt Sheehan, “Lessons From the World's Two Experiments in AI Governance,” https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035. 除此之外,《人工智能法》第二条关于适用范围的规定在一读修正案中新增 5e 款明确,《人工智能法》不适用于根据免费和开源许可提供的人工智能组件,除非它们作为高风险人工智能系统或属于其一部分投放市场或由提供者投入使用。

③ Regine Paul, *Varieties of Risk Analysis in Public Administrations: Problem-solving and Polity Policies in Europe*, Routledge, 2021.

④ 王天凡:《数字平台的“阶梯式”监管模式:以欧盟〈数字服务法〉为鉴》,载《欧洲研究》,2023 年第 2 期,第 50-77 页。

制性义务及监管措施,而非高风险人工智能系统的提供者仅需遵守一定行为准则。该首选方案被认为无论从基本价值、社会成本还是从欧盟战略需求上看,均是以最适合、最有效的方式实现《人工智能法》之目标。

但学者对这一规制路径的争议与质疑从未停止。虽然有学者认为,在监管机构面临技术创新和风险控制之间的权衡时,风险分析是理性解决问题的工具。① 但有学者针锋相对地提出,《人工智能法》草案只是表面上满足了部分"基于风险"方法所要求的理性解决问题的期望,②其"从任何科学角度上看都是不系统的",倒不如说是欧盟立法者高层对不可接受的人工智能系统范围划定的政治决定和对高风险人工智能的基于"经验法则"的分类。这种政治决定背后,更深层的是欧盟立法者选择的"话语策略",使其作为一种科学工具出现,并使之合理化,有助于将高度政治化的监管决策正当化为客观的和可信的。而这种监管决策在很大程度上是基于建立欧盟共同人工智能市场的愿景,以及欧盟立法者希望通过其话语和监管选择来实现这一愿景的意愿。③ 一言以蔽之,"基于风险"的方法并非"基于科学",而是政治决策的包装。

通过对不可接受的人工智能系统风险进行定义和禁止,欧盟立法者重申了其基本价值观,将欧盟人工智能共同市场描绘为坚定致力于基本权利保护的空间,开启"滤网效应",将不符合欧盟价值观的全球人工智能市场上的其他大型企业及其应用排除在欧盟市场之外。

在高风险领域,《人工智能法》的风险界定和监管措施在规范上均尽可能明确具体,是整个条例中规则最为集中和密集的领域,除尽可能减少侵犯个人权利的责任风险之外,更着力于提供法律确定性,为相关人工智能系统的提供者和部署者提供明确的行为要求和可预测性。

在低风险和无风险的领域,监管较为宽松,仅针对低风险有少量非强制性行为规范。欧盟立法者认为,在不可接受的风险和高风险两者已经确立严格规则的前提下,欧盟人工智能市场的"可信度"已经建立。因此,余下空间被立法者保留为"创新"和小微人工智能企业成长的空间。

---

① Tobias Krafft, Katharina Zweig and Pascal König, "How to Regulate Algorithmic Decision-making: A Framework of Regulatory Requirements for Different Applications," *Regulation & Governance*, Vol.16, Issue 1, 2022, pp.119-136.

② Olivier Borraz, "Why Regulators Assess Risk Differently: Regulatory Style, Business Organization, and the Varied Practice of Risk-based Food Safety Inspections Across the EU," *Regulation & Governance*, Vol.16, Issue 1, 2022, pp.274-292.

③ Regine Paul, "European Artificial Intelligence 'Trusted Throughout the World': Risk-based Regulation and the Fashioning of a Competitive Common AI Market," *Regulation & Governance*, 2023, DOI: 10.1111/rego.12563.

通过风险界定和分级的框架,并且允许监管机构随着人工智能用途的发展将新的应用领域纳入现有风险类别,欧盟《人工智能法》试图在法律可预测性和为人工智能发展保留空间的双重要求之间,即规则的确定性和灵活性之间保持平衡。

(二)“不可接受的风险”——禁止的人工智能系统

《人工智能法》第二章第五条规定了应被禁止的人工智能,主要包括八大类型。

针对实质性地扭曲自然人的行为,并造成或可能造成该人或他人的重大伤害的人工智能系统,条例规定了两类禁止:①第一类禁止采用超越个人意识的潜意识技术,或有目的的操纵或欺骗技术,明显损害个人或群体作出知情的决定的能力,实质性地扭曲他们的行为,从而导致其作出本来不会作出的决定;第二类禁止利用个人或特定人群因其年龄、残疾或特定社会或经济状况而存在的弱点,实质性地扭曲该人或与该群体有关的人的行为。与草案相比,议会通过版最重要的修订在于,两项禁令均不再仅适用于人工智能技术对人们造成身体或心理伤害,而是只需“重大损害”就足够了,这意味着仅造成财产损失也满足条件。并且,此类人工智能技术的提供者或部署者并不必须具有主观故意,只需损害是由人工智能操纵或剥削性做法所造成的即可。

第三类禁止主要针对“社会评分”(social scoring)人工智能系统。若此类人工智能系统在与最初生成或收集数据的背景无关的社会环境中,导致对该人或群体的有害或不利待遇,或者该有害或不利待遇,与其社会行为或其严重性不合理或不相称,则也被禁止。因为此类系统可能导致歧视性结果和排斥某些群体,侵犯了人之尊严和不受歧视的权利以及平等和公正的价值观。且依据条例,该项禁令的行为主体不再局限于“公共当局或代表公共当局”。

第四类禁止主要针对为执法目的在公共场所使用“实时”远程生物识别系统。如果生物识别数据的捕获、比较和识别都在没有显著延迟的情况下发生,此类远程生物识别系统将被禁止。② 对此,在原草案中规定了三大例外,③在一读修正案中被全部删除,但到议会的通过版中又再度“复活”,并进行了补充修正。④

除此之外,第五至第八类禁止的人工智能系统为:仅根据自然人画像(profiling)或

① 诸项禁止之共同点均在于对“将人工智能系统投放市场、提供服务或使用”的禁止,因而在各类型概括中不再列举此要件。

② 这一概念不包括用于生物识别验证的人工智能系统,如目的在于访问服务、解锁设备或安全访问场所身份验证系统。

③ 包括有针对性地搜寻特定的潜在犯罪受害者,如失踪儿童;防止恐怖袭击等对自然人生命或人身安全存在具体、实质性和迫在眉睫的威胁;以及侦查、定位、识别或起诉符合特定条件且最高刑期三年以上的刑事犯罪嫌疑人。

④ 详见下文相关内容。

对其个性特征和特点的评估,以评估或预测其实施刑事犯罪可能性的人工智能系统;①通过无针对性地从互联网或闭路电视录像中抓取面部图像,来创建或扩展面部识别数据库的人工智能系统;在工作场所和教育机构用以推断自然人的情绪的人工智能系统(出于医疗或安全原因的除外);根据自然人的生物识别数据对自然人进行分类,以推断其种族、政治观点、工会成员身份、宗教或哲学信仰、性生活或性取向的生物信息分类系统。②

可见,欧洲议会总体上对涉及伦理的复杂问题仍持保守态度。例如,对于刑事犯罪等可能导致基本权利受到严重侵犯的重要决定,欧洲议会否定了完全依赖人工智能系统,由其依据统计相关性和归纳法作出决策的可能性。

(三)高风险人工智能系统

在欧盟《人工智能法》中,监管的核心是所谓的高风险人工智能系统,条例对高风险人工智能监管的规则数量最多,内容上也是细节最密集的。无论从风险级别划定到监管措施的设置,欧盟立法者都尽力追求监管的“合比例性”。③

1.高风险人工智能系统的范围

根据《人工智能法》的规定,高风险人工智能系统的范围主要包括两个部分:其一,如果人工智能系统作为产品或产品的安全组件,为附件一所列欧盟协调立法所涵盖,且必须接受第三方合格评估,才能根据该协调立法将该产品投放市场或投入使用,则这一人工智能系统应被视为高风险人工智能系统。只要同时满足这两个条件,无论人工智能系统是否独立于上述产品投放市场或投入使用,该人工智能系统都应被视为高风险系统。其中比较重要的包括机械、玩具、无线电设备、医疗设备、体外诊断设备、车辆、民航安全、铁路系统等。

其二,条例附件三所指的人工智能系统被归属于高风险。它主要包括:(1)生物识别的系统;(2)关键基础设施;(3)教育和职业培训;(4)就业、员工管理与自营职业;(5)获得和享受基本私人服务和基本公共服务及福利;(6)执法;(7)移民、庇护和边境管制管理;(8)司法和民主进程。条例同时指出,如果人工智能系统不对自然人的健

① 此项规则被认为是欧洲议会对美国使用的 COMPAS 风险预测系统的激烈辩论的回应。条例为这一禁令设置了例外,即其不适用于支持人工评估某人是否参与犯罪活动的人工智能系统,因为人工评估已经以与犯罪活动直接相关的客观和可核实的事实为依据。

② 该禁令不包括基于生物识别数据或执法领域生物识别数据分类对合法获取的生物识别数据集(如图像)进行任何标记或过滤。

③ 篇幅所限,条例的具体条文内容并非本文重点,于此仅举要以引发更多关注和讨论,且以在一读修正案和通过版中作出重大修改或补充的核心实体法问题为例展开。大量程序性规则内容从略。

康、安全或基本权利构成重大损害风险,不对决策结果产生实质性影响,则不应被视为高风险。条例第 6 条第 3 款列明了不影响决策结果的人工智能系统适用的四种情形,作为例外的例外,附件三所涵盖的人工智能系统在对自然人进行特征分析时,应始终被视为高风险系统。条例要求欧盟委员会在咨询欧洲人工智能委员会后,不迟于 2026 年 2 月 2 日提供具体说明本条实际实施的指南,以及高风险和非高风险人工智能系统用例的实例清单。

上述范围的划定并非一成不变,满足特定条件时欧盟委员会有权通过授权法案,修改、增加或删除不再视为高风险人工智能系统的条件。但是,任何修改均不得降低欧盟对健康、安全和基本权利的总体保护水平。委员会也可以通过授权法案增加或修改附件三中所列高风险人工智能系统及其用例,只要该人工智能系统拟用于附件三所列的任何领域,并且对健康与安全构成损害风险,或对基本权利产生不利影响,且该风险等于或大于附件三中已提及的高风险人工智能系统造成的伤害或不利影响的风险。委员会应确保两类授权法案保持一致,并应考虑市场和技术的发展。

如果提供者认为附件三中提及的人工智能系统不是高风险的,则应在该系统投放市场或投入使用之前记录其评估结果,但其仍应遵守条例规定的注册义务,并应国家主管当局的要求提供评估文件。

2.对高风险人工智能系统的要求

条例对高风险人工系统提出了明确的合规要求,首先是建立、实施、记录与维护风险管理体系的义务。条例明确指出,风险管理系统应该贯穿高风险人工智能系统的整个生命周期,包括其持续迭代过程,并应当定期进行系统审查和更新。对高风险人工智能系统应进行测试,以确定最合适和最有针对性的风险管理措施,使与每种危害相关的残余风险及总体残余风险控制在可接受范围内。条例特别强调,在实施风险管理系统时,提供者应考虑高风险人工智能系统的预期目的是否可能对未成年人和其他弱势群体产生不利影响。

在数据集与数据治理方面,条例规定了相应数据集的最低质量标准。使用涉及用数据训练模型技术的高风险人工智能系统,其功能主要取决于所使用的用于训练、验证和测试的数据集。需要强调的是,议会通过版中对数据集可能存在偏见的情形做出补充规定,明确了所谓“歧视风险”。将此内容加入法案,足见欧盟立法者对算法监管的重视:包括训练、验证和测试数据集的偏见审查义务和采取措施的义务,特别是针对“反馈回路”等问题,以及例外地处理特殊类别个人数据的条件及义务。

在透明度和可解释性方面,条例规定了高风险人工智能提供者的透明度义务和向部署者提供信息的义务。高风险人工智能系统的设计和开发应确保其运行足够透明,使部署人员能够解释系统的输出并适当使用。透明度的类型和程度,应使提供者和部署者能够遵守条例规定的相关义务。条例要求高风险人工智能系统附有数字形式或其他形式的使用说明,其中包括与部署人员相关、可访问和理解的简洁、完整、正确和清晰的信息。但与先前草案一样,“可解释”这一概念仍然是极为模糊、缺乏标准的,并未明确可解释的范围和程度。

“人工监督”的规定是条例的创新点之一,“反向”体现了欧盟建立“以人为本”的人工智能环境的期望。条例要求,高风险人工智能系统的设计和开发应包括适当的人机界面工具,以便在使用期间能够受到自然人的有效监督。人工监督的措施应与高风险人工智能系统的风险、自主程度和使用环境相适应。

针对高风险人工智能系统进入欧盟市场可能呈现出的复杂样态,条例系统性地规定了价值链参与者义务。与其他数字领域的条例相比,《人工智能法》规定的监管措施总体呈现出一个重要特点:义务主体的广泛性。条例第三章的第三节专门针对各类主体规定了各自义务,虽然其主要约束对象是人工智能系统的提供者,①但义务主体涵盖了人工智能系统价值链中几乎所有的参与者。② 在某些情况下,经营者可以同时扮演一个以上的角色,因此应累计履行与这些主体相关的所有义务。例如,运营商可以同时充当分销商和进口商。并且,符合条件时主体可能发生“身份转换”。此时,最初将人工智能系统投放市场或投入使用的提供者将不再被视为该特定人工智能系统的提供者。此外,如果高风险人工智能系统是特定欧盟协调立法所涵盖产品的安全组件,则产品制造商符合条件时应被视为提供者而承担义务。

一读修正案引入了基本权利影响评估制度,并被条例最终采纳。在部署条例附件三所涵盖的高风险人工智能系统(除关键基础设施之外)之前,部署者③有义务就该系统对基本权利的潜在影响进行全面评估。一读修正案中“对基本权利的可合理预见的影响”曾引发了较多批评,因部署者面临一定的调查难度,特别是由于许多人工智

① 条例在第三章第三节专门规定关于建立质量管理体系、文件保存、自动生成日志、采取纠正措施和通知义务及与主管部门合作的义务等。

② Philipp Roos und Caspar Weitz, Hochrisiko-KI-Systeme im Kommissionsentwurf für eine KI-Verordnung, MMR 2021, 844.

③ 根据条例第 27 条,仅限于受公法管辖的机构或提供公共服务的私人实体的部署者。

能系统缺乏透明度,[①]不过这一困难由于限定在“可合理预见”的范围而得以缓和。[②]但在最终版中,这一限定却被删除,仅保留“可能对基本权利产生的影响”,对部署者而言更为严格。

根据条例规定,高风险人工智能系统在投放市场或投入使用之前应确保经过相关合规评估程序,并且在发生重大修改时,应进行重新评估。条例规定了两类合规评估程序:自评估[③]和第三方评估。第三方合规评估的范围多限于与产品有关的高风险人工智能系统,一般情况下,提供者可以选择任何公告机构进行评估。在法律适用方面,《人工智能法》对高风险人工智能系统的要求应作为相关领域立法合规要求的一部分。[④] 根据条例规定,合规评估的首要义务人是高风险人工智能系统的提供者。而实际上,部署者、授权代表及进口商等均在其各自价值链功能层面上负担相应确保合规的义务,这也意味着在义务违反时均有责任承担之可能。

条例还要求,高风险人工智能的提供者应为每个高风险人工智能系统起草一份合规声明,并在投放市场之前加贴 CE 标志,以表明其符合《人工智能法》的规定。对于纯数字高风险人工智能系统,应使用数字式 CE 标志。

此外,高风险人工智能系统的设计和开发方式应确保在整个生命周期中达到适当水平的准确性、鲁棒性和网络安全性,其技术文件应保持最新,并应从技术上允许在其生命周期内自动记录事件(日志)。并且,属于附件三所规定的高风险人工智能系统,在投放市场或投入使用之前,其提供者(或其授权代表)、公共机关或根据《数据市场法》被认定为看门人企业的部署者须在欧盟数据库中注册该系统。

(四)其他人工智能系统

对于非高风险人工智能系统,条例仅施加有限的透明度义务,包括:其一,人机交互的示明义务。条例要求,与自然人交互的人工智能系统的设计与开发中须使自然人得知其正在与人工智能系统交互,除非(对自然人而言)是显而易见的。其二,人工智

① Maximilian Frisch und Marcel Kohpeiß, KI-Verordnung, Aktueller Stand und Vergleich der Änderungsvorschläge des Rates und des Parlaments, ZD-Aktuell 2023, 01318.

② Daniel Becker und Daniel Feuerstack, Der neue Entwurf des EU-Parlaments für eine KI-Verordnung, MMR 2024, 22. 对“可合理预见”的界定,也必然带来法律上的不确定性。GDPR 第 35 条所规定的数据保护影响评估中,已经出现了与《人工智能法》所规定这一义务相类似的困难。

③ 主要针对附件 3 所列高风险人工智能系统,除用于生物识别的高风险人工智能系统需由第三方公告机构进行合规评估。

④ 因此,如果依据相关欧盟协调立法规定,产品制造商若采用了涵盖所有相关要求的协调标准,即可选择不参加第三方合格评定,则其仍须采用所有条例对高风险人工智能规定的协调标准或通用规格。例如,《人工智能法》的要求涉及确保机械安全功能的人工智能系统的安全风险,而《机械条例》(Directive 2006/42/EC)中的某些具体要求将确保人工智能系统安全地集成到整套机械中,从而不损害机械的安全。

能生成合成音频、图像、视频或文本内容的标记义务。生成式人工智能的提供者应确保其输出以机器可读格式标记,并可检测为人工生成或篡改,且应确保技术方案最新。这一特别规则是议会通过版中的新增义务。其三,情绪识别及生物识别分类系统示明义务。这些特定类别可能涉及性别、年龄、头发颜色、眼睛颜色、纹身、个人特征、种族血统、个人喜好和兴趣等方面。其四,“深度伪造”图片、音视频的披露义务。如果内容构成明显具有艺术性、创造性、讽刺性、虚构性的作品或节目的一部分,上述披露义务仅限于以不妨碍作品展示或欣赏的适当方式完成。此外,为向公众通报公共利益问题而发布的文本,若为人工智能生成或篡改,部署者亦需进行披露。除非人工智能生成的内容经过了人工审查或编辑控制的过程,并且自然人或法人对内容的发布负有编辑责任。

上述义务均不适用于经法律授权用于侦查、预防、调查和起诉刑事犯罪的人工智能系统。[①] 条例要求,上述信息最迟应在首次与自然人互动或接触时提供给该自然人。

三 争议与评价:欧盟《人工智能法》的局限性

(一)未完全消弭的争议

1.公共场所“实时”远程生物特征识别

在欧盟委员会草案发布后,关于公共场所“实时”远程生物特征识别的条款引发了普遍关注和激烈讨论。一读修正案删除了原草案规定的三种除外情形,使其成为无例外的绝对禁止。而最终通过版又再度恢复,并予以补充和修改。如此辗转反复,可见在这一问题上共识的缺乏,势必为条例的实施埋下隐患。

对于一读修正案中删除除外情形的理由,欧洲议会指出,其可能对相关人员的权利和自由具有特别的侵扰性,影响大部分人的私人生活,唤起被持续监视的感觉,还可能使公共场所生物特征识别系统的部署者拥有不可控制的权力地位,以及可能会由于技术上的不准确而导致结果偏差,并造成歧视性影响。反对删除除外情形的学者认为,完全禁止不符合比例原则,尤其是涉及某些严重威胁,如恐怖袭击、严重罪行以及追踪失踪儿童等重大公共和私人利益时。这种绝对禁止并不一定是出于基本权利保

① 唯一例外是人机交互的人工智能系统,若其可供公众举报刑事犯罪,则仍应示明。

障的目的,因为重大利益(如调查严重犯罪)可以成为限制隐私的理由。[①] 并且,在公共空间中,私人领域很少绝对受保护。事实上,一读修正案中删除除外情形的问题之所以产生争议,一方面在于此类人工智能系统可能给其部署者(通常是大型人工智能企业)带来难以控制的权力,欧盟立法者对此充满戒心;而另一方面,欧盟各国却希望在保护"基本权利""个人隐私"的名义下,仍然能有效维护自身的国家安全。[②]

条例规定,禁止为执法目的在公共场所使用"实时"远程生物识别系统,但为实现以下目标之一属绝对必要的除外:有针对性地搜寻绑架、贩卖人口或性剥削的特定受害者,以及寻找失踪人员;防止对自然人的生命或人身安全构成具体、重大和迫在眉睫的威胁,或防止真实的、现实或可预见的恐怖袭击威胁;定位或识别犯罪嫌疑人,以便对附件二所述罪行进行侦查、起诉或执行刑事处罚,且依据有关成员国的规定,该监禁或拘留的刑事处罚的最长期限须不低于四年。作为限制,决定是否使用该系统时须斟酌如果不使用将造成危害的严重性、可能性和规模,以及如果利用该制度对所有有关人员的权利和自由造成的后果,特别是这些后果的严重性、可能性和规模。使用此类系统须遵守所在成员国法律规定的保障措施和条件,完成基本权利影响评估,并在欧盟数据库中注册。[③] 条例要求成员国针对上述内容出台详细规则,并可根据欧盟法,对远程生物识别系统的使用制定更严格的法律。

值得注意的是,一读修正案曾增加一条规定禁止"事后"远程生物识别系统分析公共可访问空间的记录镜头,并且在原则禁止之下设置了例外。[④] 立法者认为其与"实时"远程系统存在不同特点和使用方式,涉及不同风险。[⑤] 但在最终通过版中,"事后"远程生物识别系统被归属于高风险人工智能系统,不再属于"禁止"之列。可见这一问题同样极具争议。条例中将这两种识别系统区别规定于不同的风险类型,却只字未提其所涉风险究竟有何不同。

2.对通用人工智能模型(General-purpose AI Models,GPAIM)的监管

---

① Daniel Becker und Daniel Feuerstack, Der neue Entwurf des EU-Parlaments für eine KI-Verordnung, MMR 2024, 22.

② Mario Martini, Gesichtserkennung im Spannungsfeld zwischen Freiheit und Sicherheit, NZVwR Extra 1-2/2022, 16.

③ 在有正当理由的紧急情况下,可在未经欧盟数据库注册或未获得授权的情况下开始使用此类系统,但不得无故拖延完成注册或申请授权。

④ 除非它们根据联邦法律受到司法授权,并且对于与《欧洲联盟运作条约》第83(1)条定义的特定严重刑事犯罪有关、为执法目的已经发生的有针对性的搜查是绝对必要的。

⑤ Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206-C9-0146/2021 — 2021/0106(COD)).

在条例的制定过程中,针对通用人工智能模型的监管是各国争论的一个关键点。法国、德国和意大利等均主张将这一技术排除在监管之外,提倡这一领域的行业自律。法、德相关部门领导人均在不同场合提及,对通用人工智能模型的烦琐监管可能会严重阻碍欧洲的创新,最终导致技术的发展绕开了欧洲大陆。针对这一问题的分歧之严重,甚至一度被怀疑可能会阻碍欧盟《人工智能法》的通过。①

在最终通过的《人工智能法》中,对通用人工智能模型的监管成为一项重点新增制度,②在第五章(第 51-56 条)中,从类型化、程序、提供者义务等方面都新增条文予以规定。依据条例,"通用人工智能模型"是指使用大规模自我监督的大量数据进行训练的人工智能模型,其具有显著的通用性,且无论以何种方式投放市场都能够胜任执行各种不同的任务,并且可以集成到各种下游系统或应用程序中,但在投放市场之前用于研究、开发或原型设计活动的人工智能模型除外。

条例首先规定了"具有系统性风险"的通用人工智能模型的两个条件:其一,根据适当的技术工具和方法(包括指标和基准),将其评估为具有高影响能力;其二,根据附件 13 中规定的标准,人工智能系统具有与前项相当的高影响能力,欧盟委员会可根据科学小组的警报或依职权主动作出决定。当通用人工智能模型用于训练的累积计算量(以"浮点计算"FLOPs 为单位)大于 $10*25$ 时,应推定其具有高影响能力。

对于符合上述条件一的通用人工智能模型,提供者应毫不拖延地通知欧盟委员会,最迟应在符合条件或知道将符合条件的两周内通知。如果委员会意识到某个通用人工智能模型存在系统性风险,而尚未收到通知,则可以决定将其指定为具有系统性风险的模型。提供者可以在通知中提出充分的证据证明:尽管该通用人工智能模型符合条件,但由于其特定特征,不存在系统性风险。如果委员会认为该异议不成立,且提供者无法证明该通用人工智能模型不存在系统性风险,则应驳回异议,该通用人工智能模型即被归入具有系统性风险之列。对于符合系统性风险认定条件二、经委员会指定为具有系统性风险的通用人工智能模型,提供者可以提出合理的重新评估的请求。提供者最早可以在委员会的指定决定后六个月提出请求,并应包含自指定决定以来出现的客观、详细和新的理由。委员会可决定根据附件 13 中规定的标准,重新评估该通用人工智能模型是否仍存在系统性风险。如果在重新评估后决定维持其指定,则提供

① Philipp Hacker und Amelie Berz, Der AI Act der Europäischen Union—Überblick, Kritik und Ausblick, ZRP 2023, 226.

② 对其监管在一读修正案中已纳入,被称为"基础模型"(Foundation Models, FM),条例最终通过版中进行了完整构建。

者最早可以在该决定做出六个月后要求重新评估。

通用人工智能模型的提供者须担负起草并更新模型的技术文件的义务以及拟定、更新,并向下游人工智能系统提供者提供信息和文件的义务。但根据免费和开放许可发布的通用人工智能模型,除非其被认定为具有系统性风险,否则其提供者可豁免此义务。除此之外,提供者还负有与委员会和主管当局合作的义务。在第三国设立的提供者还负有任命欧盟授权代表等义务。对于具有系统性风险的通用人工智能模型的提供者,条例规定其还须负担风险的评估、识别与减轻义务,跟踪、记录和严重事件的信息报告义务以及网络安全保护义务。

特别需要注意的是,条例特别明确,超大型在线平台和超大型在线搜索引擎的提供者有义务评估其服务的设计、运行和使用中产生的潜在系统性风险,包括服务中使用的算法系统的设计如何导致此类风险,以及潜在滥用引起的系统性风险。另外,此类提供者还有义务识别并减轻因传播人为生成或操纵的内容而产生的系统性风险,特别是对民主进程、公民言论和选举进程造成实际或可预见负面影响的风险,包括通过虚假信息产生的风险。这一内容虽出现于条例的鉴于条款,但确是极罕见的“点名”到具体提供者及程序的规定,足见欧盟立法者对超大型网络平台之戒心。

3.监管沙盒

人工智能监管沙盒是欧盟《人工智能法》“支持创新”的核心措施。条例所谓“监管沙盒”是指由主管机关建立的,供人工智能系统提供商或潜在提供商在监管下,根据沙盒计划在一定时间内开发、训练、验证和测试(适当情况下在真实世界中)人工智能系统的受控框架。每个成员国应在国家层面至少建立一个人工智能监管沙盒,成员国之间也可联合建立或加入现有沙盒,该沙盒最迟在条例生效之日起 24 个月内投入使用。欧洲数据保护监管机构和各成员国可以建立更多的不同级别的联合沙盒,促进跨境合作和协同发展。为避免各成员国自行其是造成规则分裂,委员会应制定实施法案,明确人工智能监管沙盒的建立、开发、实施、运营和监督的详细安排。

建立监管沙盒的目标包括:提高法律确定性,以实现合规监管;通过监管合作,支持分享最佳实践;促进创新和竞争力,推动人工智能生态系统的发展;促进基于证据的监管学习;促进和加速人工智能系统进入欧盟市场,特别是包括初创企业在内的中小企业提供的系统。

人工智能系统的提供者参与成员国或欧盟层面的人工智能监管沙盒应得到相互和统一的认可,并在整个欧盟范围内具有相同的法律效力。人工智能系统在沙盒测试

过程中发现的任何对健康、安全和基本权利的重大风险都应得到充分缓解。如果无法有效缓解,国家主管部门有权暂时或永久中止测试过程或沙盒参与,并通知欧盟人工智能办公室。对因在沙盒中进行实验而对第三方造成的任何损害,人工智能系统的潜在提供者应当根据欧盟及成员国法律承担损害赔偿责任。但最终通过版也增加了一项重要的合规免责事由:若潜在提供者遵守了具体计划的要求及参与沙盒的条款与条件,并善意地遵循了主管部门的指导,则主管部门不得以违反本条例为由对其进行罚款。

主管当局应提供在沙盒中成功开展活动的书面证明及退出报告,详细说明沙盒中开展的活动、相关结果和学习成果。人工智能系统提供者可以在合规评估或相关市场监督活动中使用此类文件证明其已遵守本条例。相关评估机构应考虑此类报告和证明,并在合理范围内加快合规评估程序。

(二)潜在困境与评价

1.基本立场:超大型平台与中小企业两重天

对全球经济竞争力的激烈争夺(伴随着各国对国家主权和安全的担忧①)在很大程度上影响着各国监管决策,并在更广泛意义上构成各国之间“人工智能竞争”的一部分。②

与先前在数字市场领域的法律(如《数字服务法》及《数字市场法》)相比,欧盟的基本立场并未改变。由于欧盟域内缺乏超大型人工智能企业,对其而言,眼前最重要的并非国际竞争的参与,而是“防御”。《人工智能法》所展现的正是如此——通过强监管进一步构建欧盟在数字科技领域的“防御工事”。欧盟立法者在规则设计时并未手下留情,而是“以攻为守”,对超大型平台进行“多面夹击”。例如,条例在鉴于条款直接“点名”超大型在线平台或超大型在线搜索引擎提供者及其具体程序,这在本条例中极为罕见,可见欧盟立法者对超大型网络平台之戒心。而对其设定的义务,除了条例明确对具有系统性风险的通用人工智能模型的义务及要求之外,还需结合《数字服务法》的规定,包括风险评估和风险缓解、独立审计等义务。并且,对超大型平台而言,应将《数字服务法》指定的机构作为执法机构。据此,一旦依据《数字服务法》被认定为超大型平台,除本条例之外,《数字服务法》等相关欧盟条例都将聚合适用、多头

① Benjamin Farrand and Helena Carrapico, “Digital Sovereignty and Taking Back Control: From Regulatory Capitalism to Regulatory Mercantilism in EU Cybersecurity,” *European Security*, Vol.31, No.3, 2022, pp.435-453.

② Regine Paul, “European Artificial Intelligence ‘Trusted throughout the World’: Risk-based Regulation and the Fashioning of a Competitive Common AI Market”.

规制。因此,超大型平台在欧盟市场的运行必须要在合规层面满足多维度的不同要求,否则将面临巨额处罚。

为了鼓励中小企业发展、促进和培育欧盟内的人工智能创新,条例在诸多方面对中小企业、初创企业设置了优惠和豁免措施:简化质量管理体系要求、支持其合规和降低其成本的规定,包括在使用监管沙盒、基本权利评估程序和第三方合规评估费用等方面的特别考虑;并且在多种情况下为中小企业和初创企业提供部署前服务,包括关于实施本条例的指导、标准化文件、认证和咨询的帮助等。欧盟在数字立法领域对于中小企业设置大量例外,自然是为鼓励发展和保护创新,而其背后更直接的考量还是扶植和培育本土新科技企业。但条例同时强调,不管是"简化"还是"帮助",并不免除其遵守高风险人工智能系统的要求或市场监督等义务。

纵观整个《人工智能法》的规则,其依然是以强制性义务设定为绝对的主题,再加上企业是否满足条件尚需经过严格而耗时的认定过程,企业将面临时间和规则不确定性的双重成本,在未被认定后还会面临处罚。因此,很难称其为对人工智能企业发展提供了宽松和促进的环境。

2.标准制定之难

欧盟通过《人工智能法》对人工智能进行定义,并划定禁止类型及高风险类型的人工智能系统,如此便将欧洲与全球人工智能市场上的其他大型参与者区分开来,并单方面将人工智能竞争空间限制在那些被欧洲标准视为道德上可接受的系统。①

但这一构建在御敌之际是否相当于自废武功?这主要体现在立法本身的定义不清,层级不定。横向立法模式之下,对人工智能的界定决定了监管调控的宽度,而包括在欧盟层面的无数次定义的尝试都揭示了明确界定人工智能一词的困难,②其中涉及大量对具体技术的描述和抽象。《人工智能法》中提到的技术和概念仍然被认为过于肤浅,无法充分反映人工智能特有的自主行动、机器学习、对数据集的依赖、缺乏透明度和不可预测性等潜在危险。不少欧盟成员国和机构在递交欧盟的意见反馈中要求欧盟立法者针对"人工智能"概念中的关键要素进一步详细说明,尤其是各种术语,如

① Regine Paul, "European Artificial Intelligence 'Trusted throughout the World': Risk-based Regulation and the Fashioning of a Competitive Common AI Market".

② Markus Kaulartz und Tom Braegelmann, Rechtshandbuch. Artificial Intelligence und Machine Learning, 1. Aufl. 2020, Kap. 1 Rn. 2 ff.

“基于知识的概念”“统计方法”等。①

对此,欧盟委员会显然深刻意识到“标准”的重要性。采取横向立法的方法,虽然在某种程度上节约了立法成本,却增加了实施成本和实施过程中的不确定性。在风险层级的划分及各自具体义务的履行方面,欧盟立法者明确将依赖标准制定机构制定具体规则。可以说,《人工智能法》的效果在很大程度上取决于后续标准的制定。因此,欧盟一方面敦促各标准制定机构按照条例规定的时间节点颁布标准;另一方面,通过条例设置的对合规人工智能系统加贴“CE”标志,暴露了欧盟委员会的野心,即让消费者、生产商和其他司法管辖区相信欧洲产品的可信度,并以此进一步推进附着在商品输出上的标准竞争。

而如何制定合理标准,其实也是对标准制定机构的重大挑战。尽管有建议将制定精确技术要求的任务委托给专业的标准制定机构,但是,标准制定历来是由行业推动的,如何确保政府和公众在谈判桌上有一个有影响力的席位将成为一个挑战。② 此外,由于欧洲缺乏大型人工智能企业,经验的缺乏以及植根于欧盟环境的用户样本量的欠缺,必然会影响对风险的评估以及详细规则的设定。若定之过细,极易导致“灵活性”的丧失,从而成为发展之阻碍;而若过粗,倘法律的确定性将受到破坏,同时带来市场主体可预测性和信心的丧失。

3.实施协调成本高

首先,由于欧盟体系本身的特点,由欧盟层面出台的条例虽然不需要各国国内法再行转化,但是其规则无论如何之具体,也无法直接细化到实际适用。为衔接到适用,除了有赖于上述标准制定之外,还会涉及欧盟规则与各成员国固有法律体系和制度如何相容的问题。而各成员国在部门法的基本逻辑、制度和概念、框架等层面都存在历史性差异;再加上立法过程中各国之间本就存在众多争议,因此,尽管随着《人工智能法》的通过,最终迎来表面上的妥协,但这些争议背后的矛盾和差异在执法层面也一定会加倍显现,所谓的“法律确定性”和“规则的统一性”可能只是一种幻觉。

其次,由于横向立法的弱点,在实施层面,如上所述,技术标准将成为《人工智能

① 参见德国联邦人工智能协会、数字波兰基金会以及德国信息和电信行业协会针对欧盟《人工智能法》草案的反馈。KI-Bundesverbands Deutschland, “Feedback to the European Commission’s Regulation Proposal on the Artificial Intelligence Act,” August 6, 2021; Digital Poland Foundation, “Feedback to the European Commission’s Regulation Proposal on the Artificial Intelligence Act,” https://digitalpoland.org/en/blog/2021/08/feedback-to-the-european-commission-s-regulation-proposal-on-the-artificial-intelligence-act; Bitkom, Positionspapier, Bitkom Principles for the Artificial Intelligence (AI) Act.

② Matt O′ Shaughnessy and Matt Sheehan, “Lessons From the World’s Two Experiments in AI Governance”.

法》的关键部分:它们需要将立法中一般规定转化为对人工智能系统的精确要求。标准一旦生效,各国法院、监管机构和技术标准机构的多年工作将转变为准确阐明《人工智能法》如何适用于不同的情况。但上文提及,欧盟这种横向立法的方法面临下列风险:负责执行监管要求的各个监管机构可能在解释和监管能力方面存在差异,从而破坏了横向立法所需的协调能力并最终危及整部法律的效力。此外,横向欧洲人工智能办公室能否有效补充各成员国和部门监管机构的能力也令人怀疑。若要在一定程度上实现条例所预设的数字单一市场目标,必须有大量实质性的干预,这不仅需要成员国及其国内各部门的配合,还意味着海量的协商成本。

此外,从发展的角度来看,若无法建立高效的沟通和协调机制,对于爆发式发展的人工智能技术而言,横向的僵化规则要么会有阻碍创新的风险,要么沦为具文。以条例关于"人工监督"的规定为例,该规定为"能够干预高风险人工智能系统的运行或中断系统,通过'停止'按钮或类似程序使系统在安全状态下停止,除非考虑到公认的最先进的技术,人为干扰会增加风险或会对性能产生负面影响"。不能否认"人工监督"可能的功能和必要性,但如何建立这一整套制度,从选任资格、知情权到实现可能性以及违反注意义务的责任,凡此种种,对欧盟及其成员国而言将会是不小的考验。申言之,以自然人的即时判断作为人工智能的"防火墙"是不是一个有效的选择?自然人在多大程度上能够真正做到?这种"停止按钮"的思路是否在用工业时代的方法来应对人工智能时代的风险?

4.市场主体合规成本增加

成本因素对市场主体而言是除了合规要求与罚则外最敏感的内容。条例自身细琐的强制性规定,叠加欧盟协调立法中的合规要求,使得企业面临"多头规制",再加上欧盟层面及成员国层面执法的协调成本,使得市场主体难以承受成本之重。尽管条例中指出,提供者可以对本条例和欧盟相关的协调立法中所要求的文件和程序等进行"集成",以确保一致性、避免重复并尽量减少额外负担。殊不知,这种"集成"本身已是成本。

欧盟立法者对此有两方面的考虑:首先,在立法方案的选择上,立法者最终选择"3+方案",认为其能够将合规成本保持在最低水平,避免因价格和合规成本上升而导致不必要的使用放缓。通过区分风险层级并分别课以相应义务,《人工智能法》希望

达到成本增加的合比例性。[①] 为说明其风险层级划分减轻了非高风险人工智能系统提供者的负担,立法者在备忘录中举例道:让用于医疗操作的人工智能与仅管理医生预约的行政活动的人工智能一样遵守《人工智能法案》中的相同要求似乎毫无意义。具体而言,对于开发或使用对公民安全或基本权利构成高风险的人工智能应用程序的企业或公共机构,遵守条例所规定的强制性义务意味着,到 2025 年,提供一项(价值)约 170000 欧元的一般高风险人工智能系统的成本为 6000 欧元至 7000 欧元。此外,对于高风险人工智能的提供者而言,验证成本(Verification Costs)可能达到 3000 欧元至 7500 欧元。对于人工智能的用户而言,根据其使用情况,每年还需要花费时间和预计 5000 欧元至 8000 欧元确保人工监督。而开发或使用任何未被归类于高风险人工智能应用的企业或公共机构所承担的成本"最多与高风险人工智能系统一样高"。[②]

其次,立法者认为,对人工智能系统提供者和用户而言,统一的标准、支持性指南和合规工具将助其遵守提案规定的要求,并最大限度地降低成本。而对运营商而言,其产生的成本与所实现的目标及其可从该提案中获得的经济和声誉利益成正比。

然而,上述成本对市场主体,尤其是中小企业而言并非小数目,对大型、超大型企业而言为避免违规处罚,实际成本可能远超上述金额。且立法者认为可能降低成本或收益增加都只是理论猜测,其或然性无法抵消必然增加的成本。

四 全球影响:人工智能监管的方案竞争

(一)域外其他国家的选择

1.美国

与欧盟形成明显反差的是美国对人工智能的规制。自 2016 年以来,美国政府极为重视人工智能领域,出台了一系列政策文件,其政策绝大多数都是以促进、推动及保持本国在人工智能领域的领导地位为主要目的。[③] 如 2016 年由白宫科技政策办公室(OSTP)下属国家科学技术委员会(NSTC)先后发布的《为人工智能的未来做好准备》

① "Proposal for a Regulation of the European Parliament and of the Council, Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts," COM(2021) 206 final, Explanatory Memorandum, 2.3.Proportionality, Brussels, 21.4.2021.

② "Proposal for a Regulation of the European Parliament and of the Council, Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts," COM(2021) 206 final, Explanatory Memorandum, 3.3. Impact Assessment, Brussels, 21.4.2021.

③ 如 2019 年由特朗普总统签发的《保持美国在人工智能领域的领导地位》的行政命令。

(Preparing for the Future of AI)及《国家人工智能研发战略计划》(The National AI R&D Strategic Plan)。前者提出七大战略,预计在非保密人工智能研究领域投入约 10 亿美元。后者历经 2019 年、2023 年两次重要修订,其核心内容并未发生改变,只是分别增加"扩大公私合作伙伴关系"战略及"协调和集中联邦政府在人工智能领域的研发投资"战略,更凸显了集中资源布局人工智能战略的思路。

2023 年 10 月,美国政府发布的《关于安全、可靠和可信地开发和使用人工智能的行政命令》(Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence),针对安全、防止偏见、公民权利保护、隐私等方面进行规定,是对人工智能领域进行监管的框架性指令。但该行政命令的内容并未直接对人工智能进行规制,仅要求联邦各部门在一定期限内采取相应措施,包括制定规则。值得注意的是,纵观该行政命令,绝大多数情况下,其所要求制定的均为"指导方针""指南和基准"或"倡议"等非强制性规范,这与欧盟《人工智能法》的直接规制性、效力的强制性形成鲜明对比。更为明显的一个特征是,相比欧盟的监管思路,该行政令"用魔法打败魔法",即注重以技术方法规制技术,而这并非仅限于"自律",而是着力于促进开发相应的监管系统和配套资源。这样一来,便达到了以产业促进的方式规制产业。

2023 年 11 月,麻省理工学院的研究人员也发布了人工智能规制白皮书,①明确其"目的是帮助增强美国在人工智能领域的广泛领导地位,同时限制新技术可能造成的危害,并鼓励探索人工智能的部署如何造福社会"。该白皮书在规制目的、策略及手段等诸多方面与前述行政令之倾向一致。白皮书的制定者们也的确关注到欧盟《人工智能法》的进程,却未对其中关于风险分配、语言模型等通用人工智能的监管等一系列内容表示赞同。

此外,美国国家标准及技术研究所(NIST)于 2019 年发布《美国人工智能领导力:联邦参与开发技术标准和相关工具的计划》,提出人工智能标准建设方案,并于 2023 年发布首个《人工智能风险管理框架》,标志着风险监管的标准建设取得进展。

2.其他国家

自 2022 年以来,英国对人工智能监控的使用呈爆炸式增长。对欧盟试图严格规制的人工智能技术,英国政府反而积极鼓励警察系统增加使用。由于英国也面临潜在的经济衰退,存在诸多经济挑战。在人工智能政策的制定上,这些不利因素可能会影

① 麻省理工学院施瓦茨曼(Schwarzman)计算学院和麻省理工学院华盛顿办事处牵头制定政策简报,并提出有关人工智能治理的建议。其中一份简报总体描述了美国的治理框架,然后其余几个主题简报,详细阐述了人工智能治理和影响的具体方面。

响监管的方式和目的。① 2023 年 3 月,英国发布《支持创新的人工智能监管方式》(A Pro-innovation Approach to AI Regulation)的白皮书,加大对人工智能的投资。2024 年 2 月 6 日,英国政府公布针对人工智能监管白皮书咨询的回应,其中“支持创新”的基本方针与最初的提案保持不变。就此观之,英国政府计划构建一个跨部门、非法定的、基于结果的人工智能监管框架,其根本目的是在创新与安全之间寻求平衡。该框架以五项核心原则为基础:安全性、保障性和鲁棒性;适当的透明度和可解释性;公平性;问责性和治理;可竞争性和补救性。英国的监管机构将通过适用现行法律和发布补充监管指南,在各自部门或领域实施该框架。英国一方面认识到进行立法的必要性,特别是针对通用人工智能系统;然而另一方面,又认为现在就进行立法还为时过早,需要更好地理解与人工智能相关的风险和挑战、监管差距以及解决这些问题的最佳方法。②

澳大利亚政府已任命一个由法律专家和科学家组成的小组,就针对人工智能研究、开发和应用的潜在“护栏”提供建议,这是该国对这一快速发展的技术进行强制监管迈出的最新一步。工业和科学部长艾德·胡锡克(Ed Husic)表示,成立该小组的目的是就人工智能的透明度、测试和责任向政府提供建议,但其已较为明显地受到欧盟立法关于风险分级的影响。③

日本政府积极推动人工智能的开发和利用,促进相关产业的增长和扩张。在 2024 财年的国家预算中,日本政府承诺投入 1640 亿日元(比上一年增加 44%)用于增强人工智能开发能力、促进人工智能采用,并为国际人工智能监管框架的形成做出贡献。有学者指出,日本的人工智能监管方法旨在利用该技术对社会的积极影响,而不是用过多的规则扼杀创新。④ 日本已明确表示将在 2024 年推动人工智能监管立法,其目的除了解决围绕人工智能的虚假信息和侵权等问题,还涉及针对基础模型的规则,包括刑事规则。⑤

(二)中国的路径选择:人工智能监管竞争?

---

① James Morales, “Japan to Regulate AI in 2024, Global Framework Fast Emerging,” https://www.ccn.com/news/japan-ai-regulation-2024-global-framework-emerging/.

② Valeria Gallo and Suchitra Nair, “The UK's Framework for AI Regulation, EMEA Centre for Regulatory Strategy,” https://www2.deloitte.com/uk/en/blog/emea-centre-for-regulatory-strategy/2024/the-uks-framework-for-ai-regulation.html.

③ 胡锡克明确表示,政策制定的目标是希望能够找到平衡,让低风险的人工智能不受阻碍地蓬勃发展。

④ Kohei Wachi and Masafumi Masuda, “A General Introduction to Artificial Intelligence Law in Japan,” *Mori Hamada & Matsumoto*, 3 January 2024.

⑤ Sarah Brady, “Japanese Ruling Party Urges Swift Action on AI legislation-Nikkei,” *Verdict*, 15 February 2024, https://www.verdict.co.uk/japanese-ruling-party-urges-swift-action-on-ai-legislation-nikkei/?cf-view.

有学者提出,现在已经是从“人工智能竞争”到“人工智能监管竞争”的时代,①为抢占规制先机,应在人工智能规制的赛道上尽早部署监管立法,以最大程度地降低风险,提高法律确定性,有助于提高各国在竞争格局中的地位。而这一论断是否成立?中国应如何处理人工智能时代的监管?

1.统一立法时机尚未成熟

以上述及,全球各主要发达国家和区域在人工智能领域的监管思路其实较为多元。欧盟目前呈现的是最严的监管,通过强制性法规直接规制人工智能系统提供者等价值链主体。相较之下,最为宽松的是英国目前的措施和计划。而美国虽然也在某种程度上开始推动通过授权立法等形式制定标准,但目前未见到直接的强制性规则,因而处于仅次于英国的中间偏宽松状态。不同国家和地区呈现出如此巨大的监管策略差异,使人不禁怀疑这种“监管竞争”是否存在,即使是存在,也一定不是规则之强制性及严厉性的竞争。

其实,如若结合各国的人工智能技术和产业发展现状,以及在此基础上各国和地区的人工智能战略,这种“区别”便容易理解了。包括欧盟在内,都明显体现出“战略”优先。不同的监管策略正是基于各国和地区在人工智能技术和产业领域的发展状况以及在全球的地位,结合各自的人工智能战略而制定的。正是因为处境不同,策略也不同。目前,其实只有欧盟具有真正意义上的人工智能立法,这与其整体在这一领域的技术和产业相较于全球其他国家不甚发达有关,对其而言,自然需要以高强度的立法和监管限制外在输入,保护内部发展,并希望世界其他各国效法其做法,跟随其规制步伐。这样一来,欧盟便可在规则的解释,甚至未来标准的制定方面均掌握稳定的话语权。美国处于人工智能领域的领导地位,并期待提供政策甚至法律的支持以保持和推进这种地位。因此,美国在人工智能监管的问题上一直“语焉不详”,这种迟疑绝非意外,对其而言,监管规则的设置不能阻碍其人工智能技术和产业的进一步发展和扩张。② 也正因其立于浪尖,对人工智能技术的风险认知也是相对而言更为充分的,因此渐渐有了标准和监管规则制定的意向。英国处于“后发”而意图“追赶”的地位,因而对人工智能的促进与推动的需求更大,并且,相较于欧盟,英国缺乏一个足够体量的内部市场和以此为基础的政策以对抗超大型境外企业。

① Nathalie A. Smuha, “From a ‘Race to AI’ to a ‘Race to AI Regulation’: Regulatory Competition for Artificial Intelligence,” *Law, Innovation and Technology*, Vol.13, No.1, 2021, pp.57-84.

② 当然,美国本土超大型人工智能企业对其政策和立法的游说和影响力亦不可忽视。

对中国而言,上述三个法域的经验皆有可借鉴之处。在人工智能领域,中国既有促进“走出去”参与国际竞争的需求,也有“防御”的必要,更有鼓励创新和培育市场的要求。因此,中国若要制定一部法律以规制人工智能,上述几个方面的法政策考量都需要权衡。虽然已有不少学者针对人工智能技术的发展所涉及的具体问题的规制进行了富有卓见的研究,①亦有对欧洲数字转型及法律规制外溢影响的相关研究,②中国信通院也于2021年公布了中国版可信人工智能白皮书,③然而,技术发展的速度可能远超研究者的预计,更不要说还需抽象形成规则并制定规范性文件,这种“滞后性”可能是难以避免的。

更为重要的是,人工智能时代才刚刚开启,通过境内外立法经验和学术研究可知,各国和地区甚至无法妥当地对“人工智能”进行界定,这一技术正在日新月异地发展和变化着,未来将会呈现何种样态及走向难以预计。若现在仓促进入统一的、强制性法规的订立,难免“出台即过时”,倘若如此,除了发挥不了预期作用之外,还可能会成为限制技术发展和创新的桎梏。是否有必要现在制定统一的《人工智能法》?还是更多地积累实践经验,④加深对各种具体问题的研究和理解,以更好地认识与人工智能相关的风险和挑战,找寻解决问题的最佳方法,且让“子弹飞一会”。退一步讲,如果认为有必要现在制定一部统一的法律防患于未然,那么规则的制定亦不宜过细,或采用过于绝对化的表述,应保留足够的空间,以增加在使用中的弹性。

2.纵横结合的监管框架

然而,暂时不制定统一的横向基本法律,并不代表在监管层面不加强法规等其他层级规范性文件的探索。达到监管的实际效果未必要以横向的统一立法为前提。更为妥当的方式可能反而是以纵向的领域规则为基础,在时机成熟时再制定横向的上位

① 参见马长山:《数字公民的身份确认及权利保障》,载《法学研究》,2023年第4期,第21-39页;高富平:《论数据持有者权 构建数据流通利用秩序的新范式》,载《中外法学》,2023年第2期,第307-327页;余凌云:《数字时代行政审批变革及法律回应》,载《比较法研究》,2023年第5期,第87-105页;张凌寒:《深度合成治理的逻辑更新与体系迭代——ChatGPT等生成型人工智能治理的中国路径》,载《法律科学(西北政府大学学报)》,2023年第3期,第38-51页。

② 徐龙第:《全球网络空间治理:核心问题、中国方案与未来方向》,载《欧洲研究》,2023年第6期,第58-79页;金晶:《欧盟的规则,全球的标准?数据跨境流动监管的“逐顶竞争”》,载《中外法学》,2023年第1期,第46-65页。

③ “White Paper on Trustworthy Artificial Intelligence,” China Academy of Information and Communications Technology and JD Explore Academy, July 2021, http://www.caict.ac.cn/english/research/whitepapers/202110/P020211014399666967457.pdf.

④ 实际上,即使在立法成本极高的欧盟,在《人工智能法》制定之前也有大量“纵向”的探索,欧洲议会在2020年也密集通过了多项涉及人工智能伦理、责任和版权方面的决议;2021年,又发布了多项关于人工智能在刑事、教育、文化和视听领域的决议。

法。中国与欧盟的体系不同,没有其所面临的成员国各自为政而导致法律碎片化并破坏数字单一市场的问题。而从美国的授权立法中可以看出,其亦指向部门立法的模式,这可能是一种更为实际的模式。

首先,纵向法规[①]与统一立法相比成本较小。纵向法规避免了统一横向立法需要面对的诸多概括和抽象困难,仅针对具体行业领域或具体人工智能类型和使用场景进行规制,规则的提取相对容易。其次,纵向法规能够更精准调控。不同垂直领域的人工智能系统开发及部署存在众多差异,精确探索的必要性显著,相应地,规则的制定亦更具有针对性。再次,纵向法规便于与时俱进地进行修改和调整,符合目前的监管在"摸石头"阶段的需要。

但是,纵向法规虽有其灵活性和精确性方面的优势,亦有须克服的弱点。因此,强有力的统一监管机构是必须的,由其在纵向法规的制定中进行规则协调,以保证法律体系的统一性。这种纵横结合的模式可能是目前最符合中国发展的选择。

3.促进以技术监管技术

在规则的执行即实施层面,鉴于现有人工智能系统日益复杂化并欠缺透明度的现状,虽然以选定自然人来履行监督职责的"人工监督"有其必要的"最终决定"意义,但若要求其"时刻保持意识到(风险)",[②]难免强"人"所难。即使将人类的理性发挥到极致,也不一定能够提供有效的监管。

因此,规则之落地与实际效果的发生,更多地通过技术进行制约可能是更好的思路。对此,或许美国的方案能够提供一些启示。而这种通过技术的制约并非仅限于"自律",而是着力于促进开发相应的监管系统和配套资源,如"红队"测试工具等。这样一来,便达到了以产业促进的方式规制产业。而这样做并未淡化作为最终决定者的人类,人工智能系统仍然仅为工具,为其设定目的,让工具发挥预期效果的最终守望者仍然是自然人。正如"黑客"与"白帽子"一般,人工智能系统在属于监管对象的同时也可以成为监管助力。

(**作者简介:**王天凡,北京航空航天大学法学院副教授;**责任编辑:**张海洋)

① 此处的"纵向立法"主要是指广义的垂直领域的规范性文件制定,并非对应个别"部门规章",而是包括必要情况下的行政法规及联合出台规章等形式。中国《生成式人工智能服务管理暂行办法》属于典型的纵向法规,即选择特定的人工智能领域,如某类算法或应用程序,制定法规以解决其在某些领域的部署问题。

② 参见欧盟《人工智能法》第14条。