

欧盟《人工智能法案》沙盒监管机制探析

曾文革¹ 曾 权²

(1. 重庆大学, 重庆 400045; 2. 阳光时代(北京)律师事务所, 北京 100032)

摘要: 欧盟《人工智能法案》通过统一立法确立人工智能沙盒机制, 推动其在欧盟范围内实施。作为实验性立法的重要实践, 该机制融合动态监管与风险预防理念, 在目标设定、参与制度、运行协调以及数据处理规则等方面体现出公平性、确定性和创新性, 并通过与 GDPR 衔接拓展人工智能测试空间。同时, 欧盟人工智能沙盒在责任豁免和防范权力寻租方面仍存在完善空间。相比之下, 中国人工智能产业快速发展, 但人工智能沙盒机制及相关实验性立法实践仍处于探索阶段。通过分析欧盟人工智能沙盒的制度设计、实验性立法特征及其不足, 可为中国人工智能沙盒机制建设提供有益借鉴。

关键词: 欧盟《人工智能法案》; 沙盒监管机制; 实验性立法

中图分类号: DF96

文献标识码: A

文章编号: 1006-1894 (2026) 05-0064-12

人工智能在金融、医疗、交通、教育等领域的广泛应用, 在推动数字化转型和提升生产效率的同时, 也带来了算法风险、数据安全以及基本权利保护等治理挑战。如何在促进创新与保障安全之间实现平衡, 成为各国人工智能治理的重要议题。目前, 全球人工智能监管逐渐形成不同路径: 欧盟通过专门立法确立风险分级监管体系, 美国、英国等主要依托行业规则、指南和责任机制进行治理。但随着人工智能技术快速发展, 传统“硬监管”与“软监管”的二元划分已难以回应治理需求, 人工智能监管正逐步向灵活、动态、协同的敏捷治理模式转变。沙盒监管机制作为敏捷治理的重要工具, 通过提供受控测试环境, 加强监管机构与企业互动, 在风险可控基础上促进技术创新。^①当前, 中国人工智能监管沙盒机制仍处于探索阶段, 相关制度在适用范围、运行程序、责任分配和监管豁免等方面尚缺乏系统设计。相比之下, 欧盟《人工智能法案》作为全球首部系统性人工智能专门立法, 在风险分级监管基础上, 于“支持创新的措施”中首次构建人工智能监管沙盒制度, 为成员国开展人工智能创新测试提供统一法律依据。研究欧盟人工智能沙盒机制, 对于完善中国人工智能动态监管体系具有重要参考价值。本文将分析欧盟人工智能沙盒机制的制度动因、设计特色及局限, 并结合中国治理需求提出完善建议。

一、欧盟人工智能沙盒机制的创设

欧盟通过统一立法方式确立人工智能监管沙盒机制, 在全球人工智能治理实践中具有较强代表性。《人工智能法案》正式确立人工智能监管沙盒制度, 并将其作为协调人工智能创新与风险治理的重要制度工具。由于人工智能监管沙盒概念在立法过程中经历了多次调整, 下文将首先梳理欧盟人工智能监管沙盒的制度内涵, 并进一步分析其创设的法律动因。

作者简介: 曾文革, 重庆大学法学院教授, 研究方向: 国际经济法学、国际环境法学; 曾权, 阳光时代(北京)律师事务所国际业务部实习律师, 研究方向: 国际经济法学。

^① 薛澜. 人工智能治理——过去是回应式治理, 现在是集中治理, 应尽快进入敏捷治理[EB/OL]. 2021, 清华大学人工智能国际治理研究院. <https://aiig.tsinghua.edu.cn/info/1296/1284.htm>. (2026-07-12).

（一）欧盟人工智能沙盒机制的定义

沙盒（sandbox）最初源于信息技术领域，主要指通过构建隔离环境对产品进行安全测试，以识别潜在风险。随着监管科技的发展，沙盒机制逐渐被应用于金融科技、能源、物联网等领域。其中，金融科技监管沙盒发展较为成熟，英国于2015年推出金融科技监管沙盒，通过提供受控测试空间，在促进创新与防范监管风险之间寻求平衡。近年来，人工智能监管沙盒逐渐成为各国（地区）探索人工智能治理的重要工具，德国、新加坡、瑞士、西班牙等国家（地区）均通过政策、试点项目或法律框架开展相关探索。^①上述实践虽然在制度名称、运行方式和监管主体方面存在差异，但均体现出通过有限范围测试实现技术创新与风险治理协调的基本理念。根据欧盟《人工智能法案》第3条第55项规定，人工智能监管沙盒是指由主管机关建立的具体、受控的框架，使人工智能系统提供者或潜在提供者能够在监管监督下，根据沙盒计划，在有限期限内开发、训练、验证和测试创新人工智能系统。该定义体现了欧盟人工智能沙盒机制以监管参与、风险控制和促进创新为核心的制度定位。

（二）欧盟人工智能沙盒创设的法律动因

人工智能监管与发展之间的平衡已成为各国（地区）治理的重要议题。过度监管可能抑制产业创新，而监管不足则可能损害个人权益、社会秩序乃至国家安全。在人工智能技术快速迭代背景下，传统静态监管模式难以适应技术发展需求，监管沙盒因此成为欧盟探索敏捷治理的重要制度工具。

1. 沙盒机制补正欧盟人工智能严格监管模式

欧盟人工智能治理长期以风险预防和基本权利保护为核心价值，并逐步形成以风险分级为基础的治理体系。《人工智能法案》通过风险分类对人工智能系统设置差异化义务，禁止不可接受风险的人工智能应用，并对不同风险等级系统及相关主体规定合规要求，同时建立高额行政罚款机制。^②此外，欧盟亦通过责任规则完善人工智能损害救济机制，2022年提出的《人工智能责任指令》曾尝试通过证据披露、过错推定以及因果关系推定降低受害者举证负担，虽然未最终形成独立立法，但其责任强化理念已通过其他制度改革得到延续。

然而，严格的风险治理模式也可能增加企业合规成本，尤其在技术快速发展和风险边界尚不明确的情况下，企业可能因难以准确判断监管要求而降低创新积极性。人工智能监管沙盒正是在此背景下成为风险监管体系的重要补充。一方面，沙盒通过提供受控测试环境，使企业能够在监管指导下识别风险、明确合规要求；另一方面，测试反馈有助于优化产品设计，降低创新过程中的责任风险。因此，人工智能监管沙盒并非弱化监管，而是在风险防控与技术创新之间实现动态平衡的重要制度工具。

2. 沙盒机制动态预防人工智能风险

随着“风险社会”的到来，现代技术发展带来的风险呈现高度复杂性与不确定性。风险规制面临的不确定性不仅来源于技术认知限制，也体现于风险概念、评价方法、科学数据以及因果关系判断等方面。因此，风险防范原则强调在技术风险尚不明确时通过技术评估与利益衡量提前采取风险控制措施。

^① OECD.Regulatory Sandbox in Artificial Intelligence Act[EB/OL].2023,OECD. https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html. (2026-07-12).

^② 《人工智能法案》第99条。

人工智能监管沙盒正是风险预防原则在人工智能治理领域的重要实践。一方面，人工智能具有影响范围广、技术复杂度高等特征以及算法歧视等潜在风险，传统事后监管模式难以及时回应；另一方面，人工智能沙盒通过为企业提供受控测试环境，使监管机关能够提前识别系统风险，推动企业优化产品设计，实现风险前置治理。此外，由于人工智能测试可能涉及个人数据和安全风险，沙盒运行仍需结合知情同意、数据保护等制度保障，实现风险控制与技术创新之间的平衡。

3. 沙盒机制有效化解信息不对称的困境

人工智能技术的发展加剧了信息不对称问题。一方面，人工智能系统具有高度复杂性，人类对其内部运行逻辑和决策机制存在认知局限，信息不对称主要存在于人工智能提供者与使用者、运营者与交互相对人以及自动决策主体与相对人之间；另一方面，监管机构与企业之间同样存在信息不对称，前者难以充分理解技术运行机制，后者则可能面临法律适用理解障碍。针对前者，欧盟《人工智能法案》通过技术文件、日志记录、使用说明等透明度义务，提高使用者和相关主体对高风险人工智能系统的理解，从而缓解信息不对称。

然而，由于人工智能具备持续学习和自我演化能力，算法黑箱问题仍难以通过透明度要求完全解决。相比之下，人工智能沙盒通过构建监管机构、企业与社会主体之间的互动平台，进一步缩小信息鸿沟。一方面，监管机构能够通过测试过程获取技术信息，提高风险识别能力；另一方面，企业能够在监管沟通中明确合规要求，降低法律适用不确定性，实现技术创新与合规治理的协调。

二、欧盟人工智能沙盒监管机制的特色与局限性

欧盟人工智能沙盒机制规定于《人工智能法案》“支持创新的措施”中，主要涉及沙盒概要、运行模式、个人数据处理、真实世界测试以及中小企业创新支持等内容（第57条至第63条），构建了欧盟层面人工智能监管沙盒的基本制度框架，并为成员国建立国家级监管沙盒提供法律依据。需要指出的是，《人工智能法案》并非首次创设监管沙盒制度，而是在吸收英国金融行为监管局（FCA）金融科技监管沙盒、英国信息专员办公室（ICO）数据监管沙盒以及欧洲相关创新监管实践经验的基础上形成。^①

随着《人工智能法案》于2024年8月1日生效，欧盟人工智能监管沙盒已由立法设计阶段进入实施阶段。下文将结合《人工智能法案》最终文本以及域外监管沙盒实践，分析欧盟人工智能监管沙盒的制度特色及其局限性。

（一）欧盟人工智能沙盒机制具体制度的特色

1. 突出对促进本土中小企业创新的目标追求

监管沙盒作为兼顾创新促进与风险治理的柔性监管工具，其目标通常包括降低创新成本、提升监管适应性以及推动新兴技术应用。不同于部分传统监管沙盒主要关注监管规则探索，欧盟《人工智能法案》进一步强化了对中小企业和初创企业创新发展的支持。根据《人工智能法案》第57条第9款，人工智能监管沙盒不仅旨在促进创新、提升竞争力、推动人工智能生态系统发展，还强调促进人工智能系统进入欧盟市场，特别是由中小型企业 and 初创企业提供的人工智能系统。

欧盟对中小企业的重点支持与其产业结构和人工智能创新生态密切相关。中小企业和初创企业

^① European Parliament. Artificial Intelligence Act and Regulatory Sandbox[EB/OL]. 2022, European Parliament. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI\(2022\)733544_EN.\(2026-07-12\)](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.(2026-07-12)).

是欧盟经济体系和技术创新的重要主体,^①但相较于大型企业,其通常面临资金、技术资源以及合规能力不足等限制。因此,《人工智能法案》通过监管沙盒降低企业参与创新监管的制度成本,为不同规模企业提供在监管指导下开发、测试和完善人工智能系统的机会。

此外,《人工智能法案》第62条进一步规定成员国应采取支持中小企业参与沙盒的具体措施,包括提供培训和指导、建立沟通渠道、根据企业规模降低参与费用,并在条件相同情况下优先支持中小企业参与测试。可见,欧盟人工智能监管沙盒并非单纯的技术测试机制,而是通过降低创新主体参与监管试验的制度门槛,实现人工智能风险治理与创新激励之间的平衡。

2. 沙盒机制在平等原则下实现差异化参与

平等原则是欧盟法的重要基本原则。《欧盟基本权利宪章》(CFR)第20条规定:“法律面前人人平等。”该原则并不排斥差异化处理,而是要求不同待遇具有客观合理性,并与立法目标保持适当联系。欧盟《人工智能法案》将平等参与作为人工智能监管沙盒的重要制度要求。根据《人工智能法案》第58条第2款b项,监管沙盒应确保“广泛、平等的参与”,并根据参与者数量、类型及需求进行调整。该要求主要体现为保障不同类型人工智能企业参与机会的平等以及在合理范围内促进中小企业和初创企业参与。

首先,监管沙盒作为实验性监管机制,有限参与范围可能造成部分企业无法进入测试,但该差异待遇并不违反平等原则。只要准入条件、测试期限等依据人工智能系统风险水平、创新价值以及合规准备程度等客观标准确定,并服务于促进创新和风险治理目标,即具有合理性。其次,《人工智能法案》针对中小企业设置优惠措施亦属于实质平等的体现。由于中小企业通常面临资金、技术资源和合规能力不足等限制,通过降低参与成本、提供指导支持以及优先参与等方式,有助于弥补其参与监管试验中的现实障碍。同时,法案依据欧盟委员会2003/361/EC号建议确定中小企业范围,确保优惠措施适用具有客观标准,避免监管优惠措施适用范围的不确定性。^②

3. 沙盒机制对欧盟《通用数据保护条例》(GDPR)进一步数据处理制度的扩展

数据、算法、算力是人工智能产业发展的三大基石。数据是继土地、人力、资本之后的第四大生产要素,关系到数字经济发展的命脉,但数据的大规模收集和处理也可能对个人隐私、安全等基本权利造成影响。因此,如何在保障数据权益的同时促进人工智能创新,成为监管沙盒制度设计的重要问题。传统金融科技监管沙盒通常将个人数据保护纳入消费者保护机制,世界银行调查显示,70%的监管机构已制定相关个人数据保障措施。^③欧盟《人工智能法案》则进一步关注人工智能监管沙盒中的数据利用问题,并通过第59条针对“为开发具有重大公共利益的人工智能系统而进行个人数据进一步处理”建立特殊规则。

在一般情况下,人工智能企业参与监管沙盒仍需遵守GDPR的数据保护要求。根据GDPR第4条,个人数据处理包括收集、记录、组织、存储、修改、使用、披露以及删除等行为。企业对个人数据的处理应具备合法性基础,包括数据主体同意、履行合同或法律义务、基于公共利益或官方授权以及合法利益等。因此,人工智能监管沙盒并非排除GDPR适用,而是在既有数据保护框架内探索创新与权益保护之间的平衡。

① 毕马威,中关村产业研究院.人工智能全域变革图景展望:跃迁点来临[EB/OL].2023,毕马威中国. <https://kpmg.com/cn/zh/insights/2023/12/prospects-for-the-global-transformation-of-artificial-intelligence.html>.(2026-07-12).

② 欧盟《人工智能法案》序言第143段、第63条。

③ World Bank.Global Experiences from Regulatory Sandboxes[EB/OL].2020,World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/912001605241080935>.(2026-07-12).

然而，GDPR 确立的目的限制原则要求个人数据进一步处理原则上应与初始收集目的保持一致。人工智能技术的发展使这一要求面临新的挑战：数据不仅来源于用户交互过程，也广泛用于模型训练、算法优化等后续环节，而后续处理目的往往难以完全对应最初收集目的。虽然 GDPR 第 6 条第 4 款、第 89 条已针对公共利益、科学研究以及统计目的的数据进一步处理设置例外，但其适用范围仍较为有限。

基于人工智能创新对数据利用的特殊需求，《人工智能法案》第 59 条进一步发展了 GDPR 的数据处理规则。对于开发医疗、环境保护、关键基础设施管理以及提升公共服务效率等重大公共利益领域的人工智能系统，符合条件的数据进一步处理可被视为与原始处理目的相符，从而降低企业因数据用途转换产生的合规障碍。但该制度并非取消数据保护要求，企业仍需采取相应保障措施，包括在独立隔离环境中开展数据处理、记录系统训练、测试和验证过程，并在沙盒参与结束或数据保存期限届满后删除相关数据。

因此，欧盟人工智能监管沙盒并非突破 GDPR 的数据保护体系，而是在风险可控的实验环境下，对个人数据进一步处理规则进行适度调整。通过在数据利用效率与个人权益保护之间建立动态平衡，欧盟既降低了人工智能企业创新过程中的数据合规成本，也维护了人工智能治理体系的可信度。

4. 在知情同意基础上拓展真实世界测试机制

真实世界测试是金融监管沙盒中的常见机制，但其能否适用于人工智能领域仍存在争议。由于模拟环境与实际应用环境存在差异，测试结果的现实适用性有限。基于提升测试有效性的考虑，欧盟《人工智能法案》将沙盒测试拓展至真实世界。根据第 3 条第 57 款，真实世界条件下的测试是指在实验室或其他模拟环境之外，为实现预期目的对人工智能系统进行的临时测试，目的在于收集可靠数据并评估系统是否符合《人工智能法案》要求。

需要强调的是，真实世界测试并非人工智能系统进入市场或获得商业部署许可，而仍属于监管机构监督下的实验性测试活动。为防止测试活动侵害个人权益，《人工智能法案》对真实世界测试设置了严格程序要求，包括申请、审查、参与者选择、测试实施、退出以及评估等环节，并要求测试始终处于主管机关监督之下。其中，知情同意原则构成重要保障，测试对象应在参与前充分了解测试性质、目的、可能影响、持续时间以及拒绝或退出权利，并基于自由、自愿和明确意思表示参与测试。^①

因此，欧盟人工智能监管沙盒通过引入真实世界测试机制，在风险可控基础上提高人工智能系统测试结果的现实适用性。一方面，该机制能够弥补模拟环境与实际应用环境之间的差异，帮助企业 and 监管机构识别潜在风险；另一方面，通过程序控制和知情同意要求，实现人工智能创新与个人权益保护之间的动态平衡。

（二）欧盟人工智能沙盒机制具体制度的局限性

1. 有限责任保护机制可能影响企业积极性

监管沙盒中的责任豁免问题本质上是实验性监管安排与现行法律规则衔接的问题，既涉及制度合法性基础，也涉及如何降低企业参与测试过程中的法律不确定性。从宏观层面看，《人工智能法案》第 57 条确立了成员国建立人工智能监管沙盒的法律义务，为监管实验提供制度基础。但从具体运行机制来看，欧盟人工智能监管沙盒的责任保护仍较为有限。以英国金融行为监管局（FCA）

^① 欧盟《人工智能法案》第61条。

金融科技监管沙盒为例，其通过限制性授权（restricted authorisation）、不强制执行函（no enforcement action letters）、个别指导（individual guidance）以及监管规则豁免等措施降低企业测试过程中的法律风险。相比之下，《人工智能法案》第 57 条第 12 款仅规定，参与主体按照沙盒计划和相关规则开展测试活动的，不因该测试行为受到相关行政处罚处罚，未进一步建立类似金融科技沙盒中的监管指导、不执行承诺等机制。

因此，欧盟人工智能监管沙盒虽通过有限责任保护降低了企业参与测试的部分风险，但其灵活性仍低于传统金融科技监管沙盒。由于人工智能企业在参与测试过程中仍可能面临商业秘密泄露、产品声誉风险以及第三方责任等问题，有限的责任保护可能影响企业参与积极性。未来，欧盟可进一步引入个案指导、不采取执法措施的监管承诺以及保险、保证等机制减轻企业合规负担，从而推动人工智能沙盒机制的成功实施。

2. 准入标准与监管机制不足可能引发监管裁量风险

监管沙盒作为由监管机关主导的实验性治理工具，其运行不可避免涉及行政裁量。合理的裁量有助于监管机关根据技术特点、风险程度以及市场需求调整监管措施，提高治理适应性；但过大的裁量空间亦可能导致准入标准不透明、监管决定缺乏可预测性等问题。

欧盟人工智能监管沙盒由成员国主管机关负责建立和运行，《人工智能法案》虽要求监管沙盒保障“广泛和平等参与”，促进人工智能创新生态发展，但对于企业具体准入标准并未进行细化规定。因此，成员国主管机关仍需结合项目情况进行综合判断。域外监管沙盒实践表明，准入通常涉及项目创新程度、消费者利益、技术成熟度、企业合规意愿以及实施能力等因素，监管机构通过综合评价决定是否准入。准入标准的不明确可能产生两方面风险：一方面，企业难以预测进入沙盒的可能性，降低制度吸引力；另一方面，监管机关较大的判断空间可能增加行政裁量风险。因此，欧盟未来应进一步明确准入标准，加强程序监督，并要求监管机关说明决定理由，以提升制度透明度。此外，监管沙盒还可能产生“监管背书”效应。市场主体可能将进入沙盒测试误认为监管机构对人工智能系统安全性的认可，企业亦可能借此进行市场宣传，从而放大监管认可效果。因此，有必要建立信息披露机制，明确沙盒参与资格不等同于产品安全认证。

从传统监管沙盒经验来看，部分金融科技监管沙盒已针对反洗钱、反商业贿赂等领域建立监督机制。^①相比较而言，《人工智能法案》对监管沙盒运行中的利益冲突防范和监督机制规定仍较有限。未来，欧盟可通过完善准入规则、强化程序监督以及建立利益冲突防范机制，降低监管裁量风险。

综上，欧盟人工智能监管沙盒通过支持中小企业参与、完善数据处理规则以及引入真实世界测试机制，实现了创新激励与风险控制之间的平衡。但其仍存在责任保护不足、监管裁量空间较大以及配套监督机制有限等问题，需要在后续实施过程中进一步完善。

三、欧盟人工智能沙盒机制独特的实施优势

实验性立法（experimental legislation）作为回应社会快速变化的治理方式，具有较长发展历史。17 世纪法国地方治理实践以及 19 世纪英国殖民地管理中均曾通过赋予地方制度调整空间开展法律实验。20 世纪美国“州立法实验室”（laboratories of democracy）理论进一步发展了该理念，即允许地方在一定权限内探索制度创新，并通过实践反馈推动法律完善。近年来，实验性立法逐渐

^① World Bank.Global Experiences from Regulatory Sandboxes[EB/OL].2020,World Bank. [https://documents.worldbank.org/en/publication/documents-reports/documentdetail/912001605241080935.\(2026-07-12\)](https://documents.worldbank.org/en/publication/documents-reports/documentdetail/912001605241080935.(2026-07-12)).

成为欧洲国家应对技术变革的重要工具，德国、法国、荷兰等国已在多个领域开展相关实践。由于人工智能具有高度复杂性和快速迭代特征，传统稳定型立法模式难以及时回应技术发展需求，欧盟《人工智能法案》遂通过监管沙盒引入实验性立法理念，在创新与风险之间寻求动态平衡。

欧盟人工智能沙盒并非试图一次性建立完整监管体系，而是通过统一立法确立基本框架，并结合成员国实践反馈持续调整规则，体现“先试验、后完善”的实验性立法逻辑。一方面，《人工智能法案》统一规定沙盒目标、基本原则和运行要求，同时将具体实施细节交由成员国主管机关探索；另一方面，法案通过评估和反馈机制总结运行经验，推动监管规则动态完善。实验性立法虽然可能面临法律确定性不足和制度协调困难等问题，但欧盟《人工智能法案》通过配套机制降低了上述风险。一方面，法案通过统一规定人工智能沙盒的基本目标、运行原则和实施要求，为成员国开展沙盒测试提供共同法律框架；另一方面，通过评估、反馈和经验总结机制，推动监管规则根据实践效果动态调整。法律实施不仅取决于规则制定，也取决于规范能否有效转化为实践制度和社会行动。因此，欧盟人工智能沙盒并非单纯的创新监管工具，而是通过“统一框架+动态反馈”的实施机制，在保持法律稳定性的同时提升人工智能治理的适应能力，体现了《人工智能法案》实验性立法模式的独特优势。

（一）实验性立法确定性的保障机制

法律的确定性可以从两个方面进行理解，一方面，法律的确定性要求规则语义清楚、概念明确，条文必须公开而不会引起社会公众的歧义；另一方面，法律必须满足公民一定时间内的心理预期，公民能合理安排自己的行为，不至于无所适从。不过，法律确定性原则绝非不可撼动，置于历史长河，任何法律都只会拥有“短期的确定性”，落后于社会实践的法律会不断被淘汰，适应发展的法律则不断更新。法律的确定性从不要求法律一成不变，而是要求法律永远不会在谁也没法预料的情况下突然发生改变。

实验性立法相较于传统立法模式，最大的挑战在于其动态调整属性。由于实验性法律通常仅确立原则性框架，并将部分规则制定权交由实施主体根据实践进一步细化，可能导致适用标准不明确和法律预期降低。欧盟人工智能沙盒机制同样面临这一问题。《人工智能法案》主要规定沙盒目标、基本原则和运行框架，具体实施规则由成员国主管机关结合实际制定，若缺乏协调机制，可能导致成员国规则差异扩大，影响制度可预期性。

对此，《人工智能法案》通过统一协调机制降低实验性立法的不确定性。一方面，法案明确将提高法律确定性作为沙盒目标之一；另一方面，《人工智能法案》第58条第1款授权欧盟委员会制定统一执行方案，对参与资格、测试程序、退出机制和评估标准等作出具体要求，为成员国沙盒运行提供统一参考。该机制既确立了欧盟层面的制度底线，又保留成员国结合自身产业特点进行制度探索的空间。因此，欧盟人工智能沙盒并非消除实验性立法的不确定性，而是通过协调规则、程序规范和评估机制，将其控制在可接受范围内。

（二）实验性立法碎片化风险的制度回应

法律协调性要求不同层级、不同类型的法律规范之间保持有效衔接，以避免规则冲突和制度碎片化，从而增强法律体系的统一性和可预测性。具体而言，法律协调性具体可表现在立法的整体性、区域协同性等方面。因此，法律协调性是确保立法有效运作、实现法治统一的重要原则。通过增强立法的整体性、协同性，提高法律规则的统一性和推动区域协同立法，能够提高法律的预测性和权威性，为社会提供行为规范指引。对于具有多层级治理结构的欧盟而言，法律协调性不仅体现为欧

盟层面规则体系内部的统一，也体现为欧盟与成员国之间监管实践的有效衔接。

实验性立法虽然能够增强法律对技术发展的适应能力，但也可能因制度探索导致实施标准不一致。人工智能监管沙盒同样面临协调性挑战。对此，欧盟主要通过统一规则框架和区域协调机制予以回应。在规则层面，《人工智能法案》要求监管沙盒在促进创新的同时建立统一运行框架，并推动最佳实践分享与跨境合作。例如，第 57 条第 5 款要求沙盒促进创新和竞争力发展，并支持参与主体之间的信息交流；第 57 条第 13 款进一步要求成员国主管机关加强跨境合作。在机构层面，《人工智能法案》设立欧盟人工智能委员会（European Artificial Intelligence Board）和人工智能办公室（AI Office），负责推动成员国执行一致性、收集实践经验并促进规则协调。^①其中，《人工智能法案》第 66 条则规定，人工智能委员会负责促进成员国主管机关执行《人工智能法案》的一致性，包括信息收集、经验交流以及统一实施实践的形成。与此同时，成员国主管机关应在人工智能委员会框架下开展合作，并向人工智能办公室和人工智能委员会报告监管沙盒的建立情况、运行经验以及相关建议，从而促进欧盟层面监管实践的持续协调。因此，欧盟人工智能沙盒并非通过完全统一化消除成员国自主性，而是在统一规则框架下，通过信息共享、经验反馈和机构协调降低制度碎片化风险。该“统一框架+成员国实施+欧盟协调”的模式，既保障了欧盟数字治理的一致性，也保留了成员国探索空间，有效回应了实验性立法可能产生的协调性不足问题。

综上，欧盟人工智能沙盒机制通过预先设置确定性保障机制和协调性保障机制，对实验性立法固有的不确定性风险进行了制度回应。这不仅提升了沙盒制度实施的可预期性，也增强了欧盟人工智能治理体系的整体协调能力，构成其人工智能监管沙盒机制的重要实施优势。

四、欧盟构建人工智能沙盒监管机制对中国的启示

在探讨中国构建人工智能沙盒监管机制的障碍与完善路径前，有必要首先分析其必要性与可行性。就必要性而言，人工智能领域法治实践尚不成熟，传统静态规制模式难以适应技术快速发展需求，应通过实验性立法提升数字治理的动态回应能力。欧盟《人工智能法案》中的沙盒机制体现了动态监管理念，通过在风险可控环境中促进技术创新，为中国完善人工智能治理体系提供了重要参考。从可行性来看，中国开展人工智能沙盒建设具备制度与产业基础。一方面，实验性立法已成为中国制度创新的重要路径。自改革开放以来，中国在经济改革、司法体制等领域积累了丰富的丰富试点经验，2015 年《立法法》第 13 条进一步为地方制度创新提供法律依据。另一方面，中国人工智能产业发展迅速，已形成涵盖基础层、技术层和应用层的产业生态。截至 2023 年，中国人工智能企业数量超过 5,700 家，占全球约 16%，为人工智能沙盒机制实施提供了产业基础。^②因此，中国构建人工智能沙盒监管机制既具有现实需求，也具备制度和产业条件。欧盟人工智能沙盒机制在立法模式、制度设计以及实施保障方面的探索，可为中国建立动态化人工智能监管体系提供借鉴。

（一）中国构建人工智能沙盒机制的立法进展

从全国层面来看，中国人工智能专门立法仍处于研究论证阶段，但学界已围绕人工智能治理规则开展持续探索。其中，《人工智能示范法》系列文本逐步纳入监管沙盒制度设计。2023 年发布的《人工智能示范法 1.0》首次提出人工智能监管试验机制，对参与条件、运行机制及责任减免作出原则

① 欧盟《人工智能法案》第64~66条。

② 毕马威,中关村产业研究院.人工智能全域变革图景展望: 跃迁点来临[EB/OL].2023,毕马威中国.<https://kpmg.com/cn/zh/insights/2023/12/prospects-for-the-global-transformation-of-artificial-intelligence.html>.(2026-07-12).

性规定；后续《人工智能示范法 2.0》《人工智能示范法 3.0》《人工智能示范法 4.0》进一步围绕风险检测、生成式人工智能以及大模型治理等问题完善相关制度。然而，目前针对人工智能监管沙盒的相关规定仍主要停留于原则层面，^①对于准入标准、测试程序、数据处理规则、责任豁免以及监管协调机制等具体内容尚未形成系统设计。

从地方实践来看，部分地区已探索具有监管沙盒特征的人工智能创新试验机制。例如，《深圳经济特区人工智能产业促进条例》提出通过测试、试验、试点等方式支持低风险人工智能产品和服务先行先试；上海亦通过人工智能产业促进相关立法推动应用场景开放和创新支持。此外，中国香港、新加坡等地区的人工智能沙盒实践也为中国制度建设提供了参考。

总体而言，中国人工智能沙盒机制尚处于探索阶段，尚未形成类似欧盟《人工智能法案》具有明确法律效力的统一制度。但现有中央立法研究、学术制度探索以及地方试点实践已为未来构建人工智能沙盒机制奠定基础。未来仍需围绕立法授权、监管主体、测试程序、责任豁免以及风险防控等方面进一步完善。

（二）中国构建人工智能沙盒监管机制的主要障碍

1. 人工智能沙盒机制立法模式定位不明

立法模式是指法律规范形成的外部表现形式。目前，人工智能立法主要存在统一与垂直立法模式以及分散立法模式两种路径，前者以欧盟、加拿大为代表，后者以英国、美国为代表。统一立法模式有助于明确监管沙盒制度基础，推动规范协调一致；分散立法模式则具有较强灵活性，但可能导致监管标准差异，增加企业合规成本。

目前，中国人工智能沙盒机制的立法定位尚不明确。从全国层面看，《人工智能示范法 4.0》虽对人工智能监管沙盒作出规定，但主要停留于原则性设计，对于运行机制、参与条件以及责任减免等具体制度尚未完善，更多发挥指导作用。从地方层面看，深圳、上海等地区已在人工智能产业促进条例中探索监管沙盒机制，呈现地方先行特点。

因此，中国人工智能沙盒机制目前呈现中央指导与地方探索并存的状态，兼具统一与分散治理特征。但立法定位不明确可能导致中央与地方规则衔接不足，不利于形成统一协调的监管体系。未来，应进一步明确人工智能沙盒机制的立法层级与制度定位。

2. 人工智能实验性立法尚处于探索阶段

人工智能监管沙盒本质上属于实验性立法工具，其有效运行依赖成熟的制度实践。然而，中国人工智能领域实验性立法仍处于探索阶段。2023 年发布的《生成式人工智能服务管理暂行办法》虽从技术治理、服务规范、监督检查以及法律责任等方面建立了生成式人工智能监管框架，但主要针对人工智能服务治理，并未涉及监管沙盒机制。此外，中国实验性立法长期以授权试点、暂行性规范和地方探索为主要形式，相较于域外实验性立法强调期限、评估和反馈机制，中国相关实践在制度评估、退出机制以及经验反馈方面仍有完善空间。目前，中国人工智能监管沙盒实践主要集中于自动驾驶等特定领域，对于生成式人工智能、通用人工智能等领域，仍主要停留于政策倡导和原则性规定阶段。《人工智能示范法 4.0》虽提出监管沙盒机制，但尚未形成关于准入条件、测试程序、数据处理以及责任安排的系统制度。因此，中国人工智能沙盒机制目前仍缺乏充分的实践基础和制度积累，实验性立法经验不足将影响其未来制度运行效果。

^① 《人工智能示范法》第89条规定。

3. 人工智能沙盒机制开展面临数据合规困境

人工智能沙盒测试涉及大量个人数据收集与处理，而人工智能数据利用具有精准化、隐秘化等特点，数据进一步处理可能导致隐私侵害、数据泄露以及算法歧视等风险。再者，鉴于数据在人工智能学习训练中的关键作用，沙盒测试对个人数据的收集和处理同样可能产生歧视性结果。

当前，中国已形成以《网络安全法》《数据安全法》《个人信息保护法》为核心的个人信息保护体系，但在人工智能场景下，如何协调数据保护与技术创新、处理模型训练与用户权益保护之间的关系仍有完善空间。人工智能沙盒兼具创新促进与风险治理双重目标，过度限制数据利用可能削弱创新功能，而过度放宽又可能损害个人权益。欧盟《人工智能法案》通过第 59 条针对开发重大公共利益人工智能系统的数据进一步处理设置特殊规则，在 GDPR 框架内实现数据保护与创新促进的平衡。欧盟这一制度值得借鉴。

相比之下，《人工智能示范法 4.0》虽已提出监管沙盒机制，但尚未对沙盒测试中的数据处理规则及创新与保护之间的协调机制作出明确规定。因此，数据合规仍是中国人工智能沙盒制度建设面临的重要障碍。

（三）欧盟人工智能沙盒机制对中国的启示

1. 推动人工智能沙盒机制中央与地方立法的协调

立法的核心在于权限配置，单一制与联邦制国家在立法模式上存在差异。欧盟通过统一立法为成员国人工智能沙盒建设提供合法性基础，并借助配套机制推动区域协调。中国作为单一制国家，实行中央统一领导与地方适度分权相结合的立法体制。若缺乏全国层面的统一制度框架，可能导致地方沙盒规则差异化。当前，中国人工智能产业已形成长三角、京津冀、粤港澳等区域发展格局，地区间产业基础和融资能力存在明显差异。^①此外，人工智能沙盒机制的创设能有效拉动投资。因此，若完全依赖地方探索，可能引发监管标准不统一及监管套利问题。

根据《立法法》第 72 条的规定，应当由全国人民代表大会及其常务委员会制定法律的事项，国务院根据全国人民代表大会及其常务委员会的授权决定先制定的行政法规，经过实践检验，制定法律的条件成熟时，国务院应当及时提请全国人民代表大会及其常务委员会制定法律。可见，《立法法》第 72 条为国务院通过行政法规开展制度试验提供了法律基础。鉴于中国人工智能产业仍处于快速发展阶段，采取“中央行政法规 + 地方实施探索”的模式更符合制度需求。具体而言，全国层面应明确沙盒机制的基本目标、运行程序、参与条件、责任安排及数据规则；地方结合产业特点开展具体实践。同时，可借鉴欧盟人工智能委员会模式，建立协调机构，负责经验收集、效果评估和区域协调，以实现统一规范与地方创新之间的平衡。

2. 借鉴欧盟人工智能沙盒建设的具体制度设计

首先，应明确促进人工智能创新，特别是支持中小企业发展的制度目标。人工智能监管沙盒兼具创新促进与风险监管功能，但欧盟《人工智能法案》特别强调降低中小企业参与成本，通过技术支持、费用减免等措施提升其参与度。中国人工智能产业同样依赖中小企业创新，目前人工智能核心企业超过 5,700 家，绝大多数属于中小企业。中小企业通常面临资金不足、融资渠道有限以及技术资源有限等问题。因此，中国可借鉴欧盟经验，将支持中小企业参与作为沙盒制度目标，并通过

^① 毕马威,中关村产业研究院.人工智能全域变革图景展望:跃迁点来临[EB/OL].2023,毕马威中国.<https://kpmg.com/cn/zh/insights/2023/12/prospects-for-the-global-transformation-of-artificial-intelligence.html>.(2026-07-12).

降低测试成本、提供技术指导等方式促进产业创新。

其次，应完善真实世界测试保障机制。欧盟人工智能沙盒将测试范围拓展至真实世界，有助于提升测试结果的应用价值，但也可能增加个人隐私和安全风险。中国未来可适度引入真实世界测试机制，并通过测试授权、知情同意、自由退出、数据保护以及重大风险预案等措施强化保障。

再次，应加强沙盒机制运行过程中的预先补正机制。由于沙盒具有实验性特征，可能面临法律确定性不足、监管裁量过大的问题。对此，我国可借鉴欧盟统一执行方案和人工智能委员会机制，建立专门协调机构，负责制定运行规则、总结地方实践经验并促进区域协调。同时，应完善反腐败、反洗钱等配套监督机制，降低权力寻租风险。

最后，应完善沙盒测试豁免机制。豁免范围和方式直接影响企业参与积极性。结合金融科技监管沙盒经验，中国可通过设置多元化责任豁免机制，在保障风险可控的基础上降低企业参与成本。同时，应强化对监管机构裁量权的程序约束，以提升沙盒运行的透明度和可预测性。

3. 强化对人工智能沙盒测试中个人数据的保护

个人数据处理是人工智能治理的核心问题。人工智能系统训练、优化和运行高度依赖数据资源，沙盒测试亦不可避免涉及个人数据收集与处理。因此，我国人工智能沙盒制度应强化与现行数据保护体系的衔接。首先，应明确沙盒测试中个人数据处理的范围和条件，要求企业遵守《网络安全法》《数据安全法》《个人信息保护法》等规定，遵循合法、正当、必要原则，防止过度收集个人信息。其次，应完善人工智能场景下数据进一步处理规则。针对人工智能模型训练和迭代过程中传统数据处理规则适用不足的问题，我国可借鉴欧盟《人工智能法案》第59条，在重大公共利益领域探索有限的的数据利用例外，并通过数据隔离、过程记录、安全保障等措施，实现数据利用与个人权益保护之间的平衡。

五、结语

欧盟通过统一立法构建人工智能监管沙盒机制，旨在推动人工智能创新发展，并在动态监管与风险预防理念下实现创新与风险治理平衡。该机制在目标设定、参与机制、数据处理规则以及真实世界测试等方面具有制度特色，但在监管裁量、防止权力寻租以及责任豁免机制等方面仍存在完善空间。作为实验性治理工具，欧盟人工智能沙盒通过成员国实施、监管机构参与以及企业和公众互动，为企业合规、公众参与和规则完善提供实践基础。针对实验性立法可能产生的法律确定性和协调性不足问题，欧盟通过人工智能委员会、统一执行方案以及成员国协调机制进行制度补正，形成“统一框架+成员国实施+欧盟协调”的治理模式，这构成其制度实施的重要优势。相比之下，中国人工智能沙盒机制仍处于探索阶段，《人工智能示范法》相关规定主要停留于原则层面，在立法模式定位、实验性立法机制以及数据合规等方面仍面临挑战。未来，中国可借鉴欧盟经验，加强中央与地方制度协调，并从支持中小企业创新、完善真实环境测试保障、强化实施协调机制、建立责任豁免制度以及完善个人数据保护规则等方面推进人工智能监管沙盒建设，以实现人工智能创新发展与风险治理的动态平衡。

参考文献：

- [1] 宾凯. 政治系统与法律系统对于技术风险的决策观察 [J]. 交大法学, 2020(1).
- [2] 布鲁诺·莱奥尼. 冯辉译. 自由与法律 [M]. 湖南: 湖南教育出版社, 2008.
- [3] 柴瑞娟. 监管沙箱的域外经验及其启示 [J]. 法学, 2017(8).
- [4] 陈光. 论我国人工智能实验性立法 [J]. 安徽师范大学学报(人文社会科学版), 2020(3).

- [5] 陈寒非. 实验性法治: 暂行立法的游击战逻辑 [J]. 清华法学, 2023(5).
- [6] 陈科霖. 狄龙规则与地方自治: 美国的实践经验及对中国的借鉴启示 [J]. 甘肃行政学院学报, 2015(2).
- [7] 陈明辉. 我国央地分权的模式与类型 [J]. 地方立法研究, 2021(4).
- [8] 程乐. 数字人本主义视域下的通用人工智能规制鉴衡 [J]. 政法论丛, 2024(3).
- [9] 范贤政. 实验性立法研究: 以广东改革开放以来的立法活动为例 [M]. 北京: 中国政法大学出版社, 2018.
- [10] 侯健新. 英国近代土地确权立法与实践 [J]. 世界历史, 2021(3).
- [11] 胡锦涛, 苏锴. 论中国语境下宪法实施和法律实施的关系 [J]. 法学论坛, 2024(1).
- [12] 胡铭, 洪涛. 我国人工智能立法的模式选择与制度展开 [J]. 西安交通大学学报 (社会科学版), 2024(3).
- [13] 华劼. 人工智能时代的隐私保护——兼论欧盟《通用数据保护条例》条款及相关新规 [J]. 兰州学刊, 2023(6).
- [14] 刘盛. 金融监管沙盒制度研究 [M]. 北京: 法律出版社, 2023.
- [15] 马长山. 数智治理的法治悖论 [J]. 东方法学, 2022(4).
- [16] 马治国, 张楠. 通用人工智能的数据风险及法治应对 [J]. 北方工业大学学报 (社会科学版), 2024(4).
- [17] 莫纪宏. 立法中的问题研究 [J]. 人大研究, 2019(5).
- [18] 彭小龙. 规范多元的法治协同: 基于构成性视角的观察 [J]. 中国法学, 2021(5).
- [19] 宋华琳. 风险规制与行政法学原理的转型 [J]. 国家行政学院学报, 2007(4).
- [20] 宋亚辉. 数字经济时代的法律挑战与学术使命 [J]. 民主与法制, 2023(48).
- [21] 汤景泰, 姚春. 计算宣传与社交媒体平台中的舆论操纵 [J]. 探索与争鸣, 2022(11).
- [22] 乌尔里希·贝克. 张文杰, 何博文译. 风险社会: 新的现代之路 [M]. 北京: 译林出版社, 2018.
- [23] 肖红军, 张丽丽, 阳镇. 欧盟数字科技伦理监管: 进展及启示 [J]. 改革, 2023(7).
- [24] 张婕龄. 实验性立法问题研究 [D]. 上海: 华东政法大学, 2022.
- [25] Bathaee, Y. The Artificial Intelligence Black Box and the Future of Intent and Causation [J]. Harvard Journal of Law & Technology, 2018, 31(2).
- [26] Buocz, T., Pfothner, S., Eisenberger, I. Regulatory Sandboxes in the AI Act: Reconciling Innovation and Safety? [J]. Law, Innovation and Technology, 2023, 15(2).
- [27] Deirdre, A. Regulators Nurturing Fintech Innovation: Global Evolution of the Regulatory Sandbox as Opportunity-Based Regulation [J]. Indian Journal of Law and Technology, 2019, 15(2).
- [28] Ranchordas, S. Experimental Legislations for AI: Sandboxes For Morals and Mores [J]. Morals + Machines, 2021, 1(1).
- [29] Truby, J., Brown, R. D., Ibrahim, I. A., et al. A Sandbox Approach to Regulating High-risk Artificial Intelligence Applications [J]. European Journal of Risk Regulation, 2022, 13(2).
- [30] Zaidan, E., Ibrahim, I. A. AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective [J]. Humanit Soc Sci Commun, 2024, 11(1).

An Analysis of the Regulatory Sandbox Mechanism of the EU Artificial Intelligence Act

ZENG Wenge ZENG Quan

Abstract: The EU Artificial Intelligence Act establishes an artificial intelligence regulatory sandbox through unified legislation, promoting its implementation across the European Union. As an important practice of experimental legislation, the EU AI regulatory sandbox integrates the concepts of dynamic regulation and the precautionary principle. It demonstrates institutional advantages in areas such as objective setting, participation mechanisms, operational coordination, and data processing rules, while expanding the scope of AI testing through coordination with the General Data Protection Regulation (GDPR). Nevertheless, the EU AI regulatory sandbox still faces challenges concerning liability exemption mechanisms and the prevention of regulatory rent-seeking. In comparison, although China's artificial intelligence industry has experienced rapid development, its AI regulatory sandbox mechanism and related practices of experimental legislation remain at an exploratory stage. By examining the institutional design, experimental legislative characteristics, and limitations of the EU AI regulatory sandbox, this paper provides insights for the future development of China's AI regulatory sandbox framework.

Keywords: EU Artificial Intelligence Act; sandbox regulatory mechanism; experimental legislation

(责任编辑: 金孝柏)