精选文章

中国、欧盟与美国的 AI 合规观察

人工智能专家斯图尔特·罗素曾提出著名的“大猩猩问题”：当人类创造出智力远超自身的物种时，是否会让自己陷入如大猩猩般被支配的境地？[1] 为了规避人工智能带来的各种危机，现实世界的立法者正在积极提供解决方案。全球正在形成以美国、中国、欧盟为首的竞争格局。根据斯坦福大学HAI发布的数据，这三大法域掌握了全球绝大多数前沿模型。[2] 由于各法域的技术发展、法律体系多有不同，从而形成了各具特色的治理范式。为了更好地理解AI领域的合规边界，本文拟对上述法域的AI合规要求进行梳理与解读，为AI行业的参与者提供参考。

一、中国

随着 2025 年 3 月 14 日《人工智能生成合成内容标识办法》的正式发布，中国人工智能产业的监管又补上了一块拼图。[3] 中国采取了纵向分层、横向分类的模式，依托上位法，针对算法推荐、深度合成、生成式 AI 等具体应用出台了专门的行政法规，构建了从数据源头到内容终端的闭环治理体系。

(一) 基础立法

《网络安全法》《数据安全法》《个人信息保

1. 斯图尔特·罗素著：《人类相容性：人工智能与控制问题》，第 132 页，2019 年版。

2. 斯坦福大学以人为本人工智能研究院（Human-Centered Artificial Intelligence, HAI）：《2025 年人工智能指数报告》（Artificial Intelligence Index Report 2025），第一章：研究与开发。

3. 国家互联网信息办公室：关于印发《人工智能生成合成内容标识办法》的通知，https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm，2025 年 3 月 14 日发布，访问日期：2025 年 12 月 09 日。

护法》一并构成中国网络法领域的顶层设计。[4]

1.《网络安全法》

2025 年 12 月，《网络安全法》迎来近十年的首次修订（以下简称《修改稿》）。《修改稿》以调整法律责任为主线，兼顾了人工智能等技术的发展趋势。[5]《修改稿》增加“人工智能”专条，包括以下四个核心内容：**第一，强化基础支撑**，明确国家支持人工智能基础理论研究及算法等关键技术的研发。这与“人工智能+”行动等政策相呼应，旨在提升模型基础能力，攻克核心技术，并推动高价值应用场景的培育。**第二，明确资源供给**，确立推进**训练数据**资源和算力等基础设施的建设。这不仅肯定了合规开展训练数据处理活动的合法性，更将训练数据视为一种关键的基础设施资源，为解决生成式 AI 的数据合规难题提供了法律指引。**第三，确定伦理规范**，将此前分散在各类文件中的伦理要求上升为法律层面的义务，强调在合规层面重视安全保障能力，确立了“伦理先行”的治理基调。**第四，完善风险治理**，确立了加强风险监测评估和安全监管的治理路径。

《修改稿》对网络运营安全和关键信息基础设施安全的法律责任作出重要修改，体现了“宽严并济，精准施策”的立法精神。本次修改，网络运行安全责任将**全面升级**，包括大幅提高罚款上限，增加对于**高管个人的处罚**，新增“关闭应用程序”等更符合移动互联网特点的处罚方式等。在全面升级处罚的同时，也为企业提供了**“避风港”**。《修改稿》对接《行政处罚法》，增加了免除、减免处罚的情形，如主动消除或减轻违法行为危害后果等，激励企业建立**事前合规体系和事中应急响应**，建立“可留痕、可举证”的完善合规体系。

2.《数据安全法》

《数据安全法》确立了**“数据分类分级”**和**“跨境传输管理”**制度，以保障国家数据主权与全生命周期的安全。生成式人工智能高度依赖大量数据进行训练和优化，《数据安全法》为人工智能奠定了数据要素治理的法律基石，其核心要求 AI 企业建立数据分类分级保护制度，严格识别并重点管控训练数据中的**“重要数据”**与**“核心数据”**。同时，强制企业落实全生命周期的数据安全管理（收集、存储、加工、销毁），确保数据来源合法、处理过程合规，并为 AI 模型开发或服务出海涉及的数据跨境传输划定了严格的安全评估与审批红线，从而确保 AI 技术在保障国家安全与公共利益的前提下进行数据的开发与利用。

《网络数据安全管理条例》为人工智能构建了覆盖“数据获取、模型训练、应用服务”的全链条合规框架[6]：**在源头**，强制要求生成式 AI 服务加强训练数据安全管理，并严禁自动化抓取（爬虫）非法侵入或干扰网络运行；**在应用端**，赋予用户对自动化推荐的“关闭权”与“标签删除权”，打破算法强制推送；**在监管端**，特别针对大型网络平台划定了利用算法实施误导、欺诈或不合理差别待遇（大数据杀熟）

4.《中华人民共和国网络安全法》（2017 年施行）、《中华人民共和国数据安全法》（2021 年施行）及《中华人民共和国个人信息保护法》（2021 年施行）。

5. 全国人民代表大会：全国人民代表大会常务委员会关于修改《中华人民共和国网络安全法》的决定，http://www.npc.gov.cn/npc/c2/c30834/202510/t20251028_449048.html，发布日期：2025 年 10 月 28 日，访问日期：2025 年 12 月 09 日

6.《网络数据安全管理条例》，2025 年 1 月 1 日起施行。

的红线，确立了技术应用必须建立在数据安全、来源合法与公平透明的基础之上。

3.《个人信息保护法》

《个人信息保护法》致力于保护个人信息权益，确立了**“告知-同意”规则**，以及对于敏感个人信息需**“单独同意”**的法律框架。而生成式人工智能在训练数据中可能涉及抓取海量个人信息，而这一过程往往是自动化的，难以在抓取瞬间区分“一般信息”或“敏感信息”，导致技术上难以达到合规标准。

(二) 行政规章

1.针对“算法推荐技术”：《互联网信息服务算法推荐管理规定》(2022) [7]

《互联网信息服务算法推荐管理规定》是中国首部针对算法推荐技术的专门性行政规章，由国家网信办等四部门联合发布，旨在通过**分级分类管理**与**算法备案制度**将技术应用纳入法治轨道。该规定明确了服务提供者的主体责任，严禁利用算法诱导沉迷、操纵榜单、实施垄断或“大数据杀熟”，同时强制打破“算法黑箱”，赋予用户对算法的知情权与拒绝权（如关闭个性化推荐），并特别强化了对未成年人、老年人及劳动者等群体的权益保护，标志着中国互联网治理正式从单纯的“内容监管”迈向了深层的“算法治理”新阶段。

2.针对“深度合成技术”：《互联网信息服务深度合成管理规定》(2023) [8]

《互联网信息服务深度合成管理规定》是针对“深度合成”的专项法规，旨在全链条监管确保技术应用的透明可控。该规定强制要求服务提供者落实实名认证与内容审核，对 AI 生成合成内容添加**显著标识**以防止公众混淆，设置人脸人声等生物信息处理的**“单独同意”**门槛，并确立了**算法备案**与**安全评估**的双重准入机制，从而有效防范利用 AI 技术进行造假、欺诈或侵害数据安全的风险。

3.针对“大模型”：《生成式人工智能服务管理暂行办法》(2023) [9]

《生成式人工智能服务管理暂行办法》是针对生成式人工智能服务的专项法规，确立了“发展与安全并重”及“包容审慎”的监管基调。该办法主要规范向中国境内公众提供的生成式 AI 服务，构建了从“源头数据治理”（要求训练数据合法、尊重知识产权与隐私）到“算法过程监管”（防歧视、算法备案）再到“终端服务规范”（内容标识、防沉迷、个人信息保护）的全链条合规体系。其核心要点在于明确服务提供者需承担“网络信息内容生产者”与“个人信息处理者”的双重法律责任，并对具有舆论属性或社会动员能力的服务实施“安全评估+算法备案”的双重准入机制，标志着中

7. 国家互联网信息办公室等:《互联网信息服务算法推荐管理规定》，2022 年 3 月 1 日起施行，https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm，发布日期：2022 年 01 月 04 日，访问日期：2025 年 12 月 09 日

8. 国家互联网信息办公室等:《互联网信息服务深度合成管理规定》，2023 年 1 月 10 日起施行，https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm，发布日期：2022 年 11 月 11 日，访问日期：2025 年 12 月 09 日

9. 国家互联网信息办公室等:《生成式人工智能服务管理暂行办法》，2023 年 8 月 15 日起施行，https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm，发布日期：2023 年 07 月 13 日，访问日期：2025 年 12 月 09 日

国 AIGC 行业正式从技术探索期迈入法治化、规范化的“持证上岗”新阶段。

4. 针对“标识”：《人工智能生成合成内容标识办法》（2025）[10]

《人工智能生成合成内容标识办法》是中国 AI 内容治理关键配套规章。该办法强制确立了“显式+隐式”双重标识制度，要求服务提供者在文本、音视频及虚拟场景中添加用户可感知的显著提示（显式），并同步嵌入包含制作要素的文件元数据（隐式）。其核心逻辑在于构建“全链条责任闭环”：上游生成者负责“打标”，中游传播平台负责“核验与提示”（区分确认为 AI、用户声明及疑似 AI 三种情形），下游应用商店负责“上架审核”，终端用户被严禁恶意篡改标识；同时，该办法兼顾了专业应用需求，允许经协议约定的用户获取无显式标识内容，但必须留存日志 6 个月。该办法旨在通过技术标准与管理规范的结合，彻底解决数字内容“人机难辨”的社会风险。

(三) 司法导向：“宽进严出”的制度供给

在立法机关、行政机关划定“红线”的同时，针对产业界最头疼的训练数据合法性问题，中国的司法机关正在探索一套更为务实的“宽进严出”裁判逻辑，以解决“海量数据需求”与“授权成本过高”之间的矛盾。

最高人民法院知识产权法庭在其官方平台上发布的《人工智能训练数据的法律风险与制度供给》一文中明确：在“输入端”**构建数据合理使用制度**，在“输出端”则采取**较为严格的制度设计**，兼顾人工智能技术发展和权利人利益保护。针对 AI 训练数据“输入端”的合法性难题，构建涵盖著作权、个人信息及企业数据的**多元化“合理使用”制度**：在著作权领域，依据“三步检验法”将 AI 训练认定为具有高公共价值的“特定且特殊情形”（如同“分子料理”般的参数重塑而非简单复制），予以**豁免**；在个人信息领域，对已公开的一般信息适用从宽解释，确立“默示同意、拒绝例外”的规则，但对个人敏感信息，需要个人的明示同意；在企业数据领域，基于其有限排他性，允许遵循公开、适当及比例原则抓取公开数据，前提是不得破坏技术防护措施或在“输出端”损害权利人的核心竞争优势。[11]

通过对输出结果的严管，倒逼企业在技术层面（如过滤机制、对齐训练）提升治理能力，而不是在数据获取阶段因噎废食。这意味着，中国未来的司法裁判可能不会简单地判决“抓取即非法”，而是更看重“如何使用数据”以及“生成了什么结果”。对于企业而言，这提供了一个重要的合规抗辩思路：**证明数据抓取的必要性，并展示对生成结果的有效管控能力**。

纵观全局，中国已构建起一套“纵向分层、横向分类、全链条覆盖”的监管体系。以“三法”确立数据主权与权益红线；以算法推荐、深度合成、生成式人工智能等规章实现对各个具体技术场景的治理；在执行端，则通过双重准入机制（算法备案+安全评估）与最新的“显

10 国家互联网信息办公室等:《人工智能生成合成内容标识办法》, 2025 年 9 月 1 日起施行, https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm，发布日期：2025 年 03 月 14 日，访问日期：2025 年 12 月 09 日

11 最高人民法院知识产权法庭，亓蕾法官:《人工智能训练数据的法律风险与制度供给》，https://mp.weixin.qq.com/s/x6fZcFx7DUrM_2T_EH592w，发布日期：2025 年 12 月 04 日，访问日期：2025 年 12 月 09 日

隐性标识”制度，对 AI 从训练数据、算法逻辑到生成内容进行穿透式监管，形成一套以“安全可控”为核心，试图在防范风险的同时为技术产业化划定清晰跑道的“中国方案”。

二、欧盟

欧盟作为全球合规监管高地，率先推出了《欧盟人工智能法案》（EU AI Act，以下简称“AI 法案”）。[12] AI 法案采取了“风险分级”的原则，分为：禁止类、高风险、有限风险和最小风险。法案首先划定了红线（第 5 条），禁止潜意识操纵、社会信用评分及公共场所实时远程生物识别等具有不可接受风险的行为。违者将面临最高 3500 万欧元或全球营收 7% 的罚款。

对于涉及基建、就业、司法等领域的“高风险 AI”，企业需履行类似产品上市前的全流程合规义务，特别是第 13 条的透明度要求与第 14 条的“人在环路”（Human-in-the-loop）机制，强制保留人类对算法的干预权。此外，第 5 章为通用目的人工智能模型（GPAI）制定了协调规则，特别是针对具有系统性风险能力的模型规定了分类标准和对模型提供者的信息、文档及评估义务。

在数据要素层面，《通用数据保护条例》（GDPR）与《数据法案》（Data Act）确立了隐私保护与工业数据共享的边界；[13] 在平台载体层面，《数字服务法案》（DSA）与《数字市场法案》（DMA）对算法推荐透明度及科技巨头的垄断行为实施穿透式监管；[14] 在责任兜底层面，《产品责任指令》更是将 AI 软件正式纳入“产品”范畴，确立了严格的无过错责任原则。这些法律与《AI 法案》互为咬合，共同构成了欧盟对人工智能从研发、运行到赔偿的全生命周期治理体系。[15]

三、美国

2025 年 12 月 11 日，特朗普签署行政令，旨在为人工智能建立一个负担最小的国家标准，而非五十个不协调的州标准。行政令中指出，目前各州各自为政的监管模式加大了企业的合规难度、部分州法强制企业在模型中植入

12. 欧洲议会和理事会第(EU)2024/1689 号条例（《人工智能法案》），2024 年通过，并将分阶段施行

13. 《通用数据保护条例》(General Data Protection Regulation, GDPR)。欧盟立法，于 2018 年 5 月施行，旨在加强和统一所有欧盟公民的个人数据保护，是数据要素中个人隐私保护的法律基础。《数据法案》 (The Data Act)，2025 年 9 月 12 日施行，旨在通过规范物联网（IoT）设备产生的工业数据的访问和使用权，促进非个人数据的共享，是工业数据要素流通的关键法律框架。

14. 《数字服务法案》(Digital Services Act, DSA)。欧盟法规，2024 年 1 月 1 日施行，旨在统一在线平台（包括社交媒体、电商和托管服务）的责任规则。重点关注算法推荐的透明度、内容审核流程、打击非法内容及对超大型在线平台（VLOPs）实施额外监管义务。《数字市场法案》(Digital Markets Act, DMA)。欧盟法规，2024 年 3 月 6 日施行，旨在规范在数字领域拥有显著市场影响力的大型在线平台的行为。它通过列出一系列禁止和强制性行为，遏制科技巨头的垄断行为，促进公平竞争

15. 《产品责任指令》(Product Liability Directive, PLD)。欧盟指令，2024 年 12 月 8 日施行，旨在统一成员国关于有缺陷产品责任的法律。该指令经过修订后，明确将 AI 系统和软件纳入“产品”的定义范围，确立了当有缺陷的产品（包括 AI）造成损害时，制造商承担严格的无过错责任的原则。

意识形态偏见以及不当干预州际贸易活动。[16]

该行政令指示司法部对与该行政令相冲突的州人工智能法律提起诉讼；授权商务部对于实施过重 AI 监管规则的州暂停联邦资金。该行政令直接批评科罗拉多州的《人工智能法案》禁止算法歧视的条款变相强迫模型修改真实输出。鉴于行政令的合宪性将面临各州总检察长的激烈挑战，在法院正式作出裁决之前，各州法律仍然有效。

由此可见，美国 AI 合规的主要问题已由各州标准不一导致合规成本过高转变为联邦与州的规则相斥。联邦政府不仅允许企业不遵守州法，甚至通过政府采购条款禁止企业实施部分原本认为是合规的行为。企业将陷入“合规二选一”的困境——为了符合加州、科罗拉多州或欧盟的法律而进行的算法偏见审计、反歧视影响评估等措施，可能反而违反了联邦的要求，导致丧失联邦政府采购的竞标资格。[17]

由于诉讼结果的不确定性，建议企业不要拆除现有的合规架构，以免面临州政府的追溯性执法。同时，仍然坚持最大公约数的策略，但做好版本分叉准备：对于底层模型训练、数据治理和安全测试仍应以欧盟《人工智能法案》和加州标准为基准，并在推理层实施动态对齐架构，以应对不同的合规要求。同时，尽管联邦层面在“去监管”，但联邦贸易委员会的反欺诈执法权并未被剥夺，建议重点记录算法决策过程，以应对州级法律挑战。

四、总结与展望

面对上述复杂的法律规定，企业必须摒弃“先发展、后合规”的方式，而是随着产品开发构建一套内生性的合规系统。首先，应设立跨部门的 AI 治理委员会，统筹法律、技术与伦理风险。其次，在产品出海前必须完成自由实施（FTO）检索，系统排查模型架构与训练数据的知识产权等法律风险。更重要的是，企业需践行“设计即合规”理念，从数据采集阶段的合法性基础，到开发阶段的偏见测试，再到部署阶段的标识溯源，建立完整的合规证据链。同时，企业还需警惕出口管制与制裁风险，建立实时的用户筛查机制，防止因向受制裁实体提供算力服务造成不利的法律后果。

未来的核心竞争力，将不仅仅取决于数据与算力，更取决于系统的透明度、稳健性与法律责任的可问责性。唯有在清晰、稳健的规则护航下，人工智能方能真正行稳致远。

16. 《消除各州法律对国家人工智能政策阻碍》，白宫，https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/，发布日期：2025 年 12 月 11 日，访问日期：2025 年 12 月 16 日。

17.《防止联邦政府出现“觉醒”人工智能》行政令 (Executive Order on Preventing Woke AI in the Federal Government)，白宫，https://www.whitehouse.gov/presidential-actions/2025/07/preventing-woke-ai-in-the-federal-government/ 发布日期：2025 年 7 月 23 日，访问日期：2025 年 12 月 16 日。该行政令确立了将觉醒 AI 排除在联邦政府的采购范围之外的规则。

附表：中国、欧盟、美国 AI 监管框架核心要素对比（2025 年）

维度	中国	欧盟	美国
域外效力	● 强。适用于向中国境内公众提供服务的公司。	● 强。适用于向欧盟市场投放AI系统的提供者，或其输出结果影响到欧盟境内个人的提供者。	● 加州SB53（前沿模型法案）适用于向加州市场投放模型的提供者。 ● 加州SB942（透明度法案）适用于在加州可访问且月用户超过100万的AI系统提供者。
事前准入	● 强制算法备案和安全评估：具有舆论属性或社会动员能力的算法推荐服务、提供深度合成服务、关键信息基础设施运营者和处理达规定数量个人信息的处理者。	● 不可接受风险的AI被禁止。 ● 高风险AI系统强制评估，必须在投放市场前进行强制合格评定。 ● 具有系统性风险的通用AI模型（GPAI）必须向欧盟AI办公室通报并承担额外义务。	● 州级强制披露：加州SB53要求大型前沿AI开发者在模型发布前发布通用安全框架和透明度报告，说明风险管理和网络安全措施。 ● 科罗拉多州法律要求高风险系统部署者实施考虑AI RMF的风险管理政策。
全球共性的关键合规动作	● 组织设立：建立内部管理和问责框架：定义角色职责，制定内部政策，并要求工作人员具备操作AI系统的必要知识（AI素养或培训）。 ● 开发与数据：明确目的和范围，并确保数据质量：在系统设计时定义其预期目的，识别潜在风险，并实施数据治理措施，确保用于训练/测试的数据集满足质量和代表性要求。 ● 评估与测试：进行前瞻性风险评估和性能测试：在部署前对系统进行充分测试，评估其准确性、鲁棒性和潜在风险，特别是对个人权利和安全的影响。 ● 运营与问责：保持透明度，确保人工可解释性和可干预性，并进行事故响应：确保用户了解其在与AI系统交互，并提供机制以便人工进行监督和干预。在发生安全事件时，须及时报告和采取补救措施。		

维度	中国	欧盟	美国
关键合规动作区域特性	● 完成算法备案、安全评估、伦理审查。	● 判断所属风险类别 ● 建立技术文档、使用说明与人类监督 ● 任命欧盟授权代表	● 执行NIST RMF框架 ● 纽约市：招聘算法需第三方审计并公示
处罚力度	● 针对个人信息，最高可处5000万元或上一年度营业额的5%，并可暂停相关业务、吊销营业执照。 ● 处罚高管个人，最高100万元人民币；处罚公司，最高1000万人民币。	● 禁止类业务最高3500万欧元或全球营收7%，高风险业务最高1500万欧元或全球营收3%罚款。 ● 针对个人数据，GDPR最高可处2000万欧元或全球营业额的4%。	● FTC以不公平或欺诈行为可销毁算法。 ● 加州：前沿模型违规最高100万美元罚款。 ● 刑事犯罪可处以罚款和最高3年监禁。

邮箱:LTBJ@lungtin.com 网站：www.lungtin.com

关于该文章，如需了解更详细的信息，请与本文作者联系。

周子琪

律师

周子琪女士擅长处理知识产权相关法律业务，对专利无效、专利侵权、技术秘密侵权诉讼等业务具有丰富的经验。