CSA GCR cloud security alliance® GREATER CHINA REGION

CSA cloud security alliance®

中美欧人工智能治理的典型法律政策比较与通用风险治理框架适用（2024年度观察）

隐私与个人信息保护法律工作组的官网地址是：

https://c-csa.cn/mobile/research/union-detail/i-1610.html

联盟简介

云安全联盟（Cloud Security Alliance,CSA）是中立、权威的全球性非营利产业组织，于2009年正式成立，致力于定义和提高业界对云计算和下一代数字技术安全最佳实践的认识，推动数字安全产业全面发展。

云安全联盟大中华区（Cloud Security Alliance Greater China Region, CSA GCR）作为CSA全球四大区之一，2016年在香港独立注册，于2021年在中国登记注册，是网络安全领域首家在中国境内注册备案的国际NGO，旨在立足中国，连接全球，推动大中华区数字安全技术标准与产业的发展及国际合作。

我们的工作

联盟会刊下载地址
了解联盟更多信息

加入我们

CSA大中华区官网
(https://c-csa.cn)

点击会员

加入联盟

填写相关申请信息

成为CSA会员

JOIN US

目 录

致谢

《中美欧人工智能治理的典型法律政策比较与通用风险治理框架适用》由隐私与个人信息保护法律工作组专家编写。感谢以下专家和单位的贡献：

组长：

原浩

参编专家：

江翔宇　　史宇航　　邢海韬

审校人员及其他贡献者：

方婷　　张元恺　　赵晨曦　　贺志生

研究协调员：

高健凯

感谢以下单位的支持与贡献：

北京天融信网络安全技术有限公司　　华为技术有限公司

（以上排名不分先后）

关于研究工作组的更多介绍，请在 CSA 大中华区官网（https://c-csa.cn/research/） 上查看。

在此感谢以上专家及单位。如报告有不妥当之处，敬请读者联系CSA GCR秘书处给予雅正！ 联系邮箱research@c-csa.cn；国际云安全联盟CSA公众号

序言

在数字化浪潮席卷全球的今天，人工智能（AI）作为推动社会进步和经济发展的关键技术，正以前所未有的速度和规模渗透到我们生活的方方面面。从智能制造到智慧医疗，从金融科技到教育革新，AI 的应用正逐步重塑着我们的世界。

对当前的人工智能热潮是否和采取何种监管，各主要国家和地区的政策法律有共性之处，也有各自的地缘、文化、哲学伦理等因素相互作用而产生的差异。互有参照、借鉴的原则、方法和进路比较，对消除智能鸿沟、提升全球人工智能的治理水平基准无疑具有重要意义，这也是 CSA 作为中立的国际组织，在跨国跨区域的法治、技术对比中寻求治理共识和实现技术普惠的初衷。

本报告《中美欧人工智能治理的典型法律政策比较与通用风险治理框架适用》应运而生，旨在深入分析和比较中国、美国和欧盟在人工智能治理领域的典型法律政策，探讨其在不同文化、经济和政治背景下的异同，并尝试构建一个通用的风险治理框架。报告不仅聚焦于当前的立法成果，更着眼于未来发展趋势，以期为全球人工智能的健康发展提供参考和启示。

报告跟踪对比了中美欧的年度主要政策法律进展，从风险管理的思路出发通过设定“可比较项”进行了纵向比较，同时通过风险提示和控制措施映射，“回归”到中国新近出台的合规框架，使得对人工智能的安全治理可以实现一条从政策法律要求，到风险控制措施规制的可行路径。

随着技术的不断演进，人工智能治理的法律政策也将随之更新。我们期待本报告能够激发更多的讨论和思考，共同推动构建一个更加公正、透明、可持续的人工智能治理体系。让我们携手前行，在确保安全和伦理的前提下，充分利用人工智能的潜力，为全人类的福祉贡献力量。

李雨航 Yale Li

CSA 大中华区主席兼研究院院长

一、摘要

如以 1956 年达特茅斯学院研究会议为标志，人工智能至今大致有过三次快速发展的热潮。在正处于第三次热潮的当下，审视既有的人工智能“历史遗留问题”，应对硬件架构、算力和模型算法的深刻变革带来的“新问题”，需要从不同国家、地区的阶段性、区域性和行业性的立法、政策和规则中吸取智慧。

本报告聚焦中美欧截至 2024 年以来对人工智能进行阶段性立法、政策的成果，通过设定不同的风险域进行细致比对，力图呈现典型国家对人工智能治理的典型做法，可以作为了解全球人工智能监管态势的起点。同时也需要注意，一方面随着技术迭代，对人工智能的法律、伦理治理仍处于快速变化的进程中，对风险的认识和控制随时可能会发生较大程度的更新；另外，目前已经有近百国家、地区开始至少围绕人工智能的某些方面进行不同程度的监管（或治理），但联合国最新《为人类治理人工智能》文件又显示在 193 个会员国中，只有 7 个国家参与了近期提出的重要人工智能治理举措，118 个会员国完全缺席1。这说明即使对人工智能的治理本身，认识上也存在极大分歧。因此读者在本报告基础上，亦应进一步了解其他国家、地区，以及包括联合国在内的国际组织在人工智能治理方面的努力和成果，以形成对人工智能国际治理全貌的自行判断和整体认识。

本报告主要涉及两方面内容，一是对中美欧典型国家、地区的人工智能立法政策进行阶段性汇总与比较分析，力图呈现不同监管关注和其形成背景；二是结合全国网络安全标准化技术委员会 2024 年 9 月发布的《人工智能安全治理框架》——该框架可视为是对目前我国人工智能技术、模型发展中的风险控制观念的集大成，将各国主要法律政策监管的要求与之进行框架层面的关联，并寻求建立两者之间的一种映射，以便于人工智能的从业者可以快速实现对监管规则的整体判断和匹配，为相关活动的开展提供符合合规要求的依据参考。

二、中国人工智能法律政策

2017 年 7 月 8 日，国务院印发《新一代人工智能发展规划》（以下简称“《规划》”），系我国首次以顶层规划的形式布局人工智能发展战略，明确了中国人

1 https://news.un.org/zh/story/2024/09/1131551

工智能技术发展的战略目标与重点任务，标志着人工智能正式成为国家战略层面的重点。《规划》提出了“三步走”的战略目标，明确了 2020 年、2025 年和 2030 年三个时间节点及分步战略，包括人工智能理论技术创新、产业经济发展、科技人才培养以及法律伦理规范等方面，旨在实现人工智能行业的持续健康发展，保证我国在人工智能领域的国际竞争力。

在《规划》的布局与指导下，我国对人工智能的发展高度重视，从人工智能发展的法律及伦理规范、重点政策支持、技术标准化和知识产权保护以及安全监管和评估等层面颁布了一系列法律政策，已然构建了较为全面的人工智能领域发展与监管框架。此外，人工智能技术的发展也离不开海量数据的投入。自 2017 年《网络安全法》、2021 年《个人信息保护法》和《数据安全法》先后出台以来，我国逐渐形成了以“数据三法”[2]为核心的数据流通利用监管体系，为人工智能发展的数据底座安全提供了法律保障。

目前，我国人工智能领域立法呈现出“从特殊到一般”的特点。一方面，随着人工智能技术的发展与应用，我国近年来针对性地出台了包括《互联网信息服务算法推荐管理规定》、《互联网信息服务深度合成管理规定》等在内的一系列专门性规范，其中，2023 年国家网信办联合七部门共同发布的《生成式人工智能服务管理暂行办法》（“《暂行办法》”）更是全球范围内领先的对生成式人工智能进行规制的立法实践。《暂行办法》对生成式人工智能服务提供者的行为进行规范，并在责任设置方面进行了一定程度的限缩，是我国现阶段关注和促进人工智能技术健康发展和规范应用的集中体现。此外，2024 年以来针对生成式人工智能训练数据合法合规性、典型行业应用等问题进行了研究，后续亦可能纳入出台的顶层设计。

另一方面，我国已将人工智能领域的立法纳入立法工作计划。根据国务院发布的年度立法工作计划，2023 年、2024 年人工智能法草案连续两年列入国务院预备提请全国人大常委会审议项目。全国人大常委会于 2023 年 9 月发布的“十四届全国人大常委会立法规划”中，将“推进科技创新和人工智能健康发展”的立法工作列入“第一类立法项目”，即“条件比较成熟、任期内拟提请审议”。在全国人大常委会公布的 2024 年度立法工作计划中，亦重申“网络治理和人工

[2] 指《网络安全法》《数据安全法》和《个人信息保护法》。

智能健康发展等方面的立法项目，由有关方面抓紧开展调研和起草工作，视情安排审议”。

2024 年 7 月，中国人大常委会法制工作委员会进一步阐述了人工智能立法的思路3：一是优先考虑灵活适用现有法律规则，通过法律解释或司法解释，解决人工智能发展过程中面临的突出法律问题，比如涉及大模型训练数据的合法合规、知识产权合理使用规则针对人工智能的适用等方面。二是对于某些人工智能应用的具体场景，可以通过授权立法的方式，让地方在立法权限范围内先行先试，或者参考目前智能网联汽车领域的试点做法，由国家有关部门组织示范应用，开展试点。三是针对影响产业发展的痛点难点问题，在迫切需要法律予以规范的领域，坚持“小快灵”立法原则，通过修改现行法律的方式解决。此外，人工智能立法也受到了中国法学界的高度重视，2024 年以来有学者发布《人工智能法（学者建议稿）》、《人工智能法示范法》等，为我国今后的人工智能正式立法提供了一定参考。

以下基于《规划》的思路和近年来我国在人工智能领域出台的主要法律政策，从四个方面简要分析归纳：

（一）法律及伦理规范

人工智能，伦理先行。2018 年 10 月 30 日，习近平总书记在主持中共中央政治局第九次集体学习时，指出要“加强人工智能相关法律、伦理、社会问题研究，建立健全保障人工智能健康发展的法律法规、制度体系、伦理道德”。2024 年 7 月 4 日世界人工智能大会暨人工智能全球治理高级别会议上发表的《人工智能全球治理上海宣言》亦明确，我们要“推动制定和采纳具有广泛国际共识的人工智能的伦理指南与规范，引导人工智能技术的健康发展，防止其被误用、滥用或恶用”。目前，我国在人工智能技术伦理领域出台的相关政策及规范主要包括以下内容（见表格），其中 2023 年颁布的《科技伦理审查办法（试行）》是我国科技伦理规制的最新实践，对从事人工智能等科技活动的单位提出了具体的伦理审查要求。

3 《中国人工智能立法怎么走？》，https://new.qq.com/rain/a/20240711A09I9000

表 1　我国在人工智能技术伦理领域出台的相关政策及规范

名称	发布机构	发布时间
《新一代人工智能治理原则——发展负责任的人工智能》	国家新一代人工智能治理专业委员会	2019 年 6 月 17 日
《网络安全标准实践指南—人工智能伦理安全风险防范指引》	全国信息安全标准化技术委员会	2021 年 1 月 5 日
《新一代人工智能伦理规范》	国家新一代人工智能治理专业委员会	2021 年 9 月 25 日
《关于加强科技伦理治理的意见》	中共中央办公厅、国务院办公厅	2022 年 3 月 20 日
《科技伦理审查办法（试行）》	科技部等十部门	2023 年 9 月 7 日

（二）重点政策支持

《规划》指出，要完善支持人工智能发展的重点政策。具体包括财税优惠、数据开放等一系列适应人工智能行业发展的政策体系，这些政策一方面能够为人工智能技术的发展提供有力保障，另一方面对人工智能发展所可能带来的社会问题能够实现有效应对。随着人工智能技术的不断发展。目前，我国出台的对人工智能发展的支持政策涉及税收、教育等方面，主要包括如下内容。

表 2　我国出台的对人工智能发展的支持政策

名称	发布机构	发布时间
《促进新一代人工智能产业发展三年行动计划（2018-2020 年）》	工信部	2017 年 12 月 13 日
《高等学校人工智能创新行动计划》	教育部	2018 年 4 月 2 日
《新一代人工智能产业创新重点任务揭榜工作方案》	工信部	2018 年 11 月 8 日
《国家新一代人工智能创新发展试验区建设工作指引》	科技部	2019 年 8 月 29 日
《关于“双一流”建设高校促进学科融合加快人工智能领域研究生培养的若干意见》	教育部、发改委、财政部	2020 年 1 月 21 日
《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》	科技部等六部门	2022 年 7 月 29 日
《关于支持建设新一代人工智能示范应用场景的通知》	科技部	2022 年 8 月 12 日
《人形机器人创新发展指导意见》	工信部	2023 年 10 月 20 日
《关于加快传统制造业转型升级的指导意见》	工信部等八部门	2023 年 12 月 28 日

《关于推动未来产业创新发展的实施意见》	工信部等七部门	2024 年 1 月 18 日

（三）技术标准化和知识产权保护

人工智能的发展离不开技术标准的统一与知识产权的保护。《规划》指出，要“鼓励人工智能企业参与或主导制定国际标准，以技术标准‘走出去’带动人工智能产品和服务在海外推广应用。加强人工智能领域的知识产权保护，健全人工智能领域技术创新、专利保护与标准化互动支撑机制，促进人工智能创新成果的知识产权化”。目前我国正不断完善人工智能领域的技术标准，并通过司法实践探索人工智能技术相关知识产权保护规则。其中，关于人工智能领域的技术标准，我国主要颁布了以下政策及标准，其中近期发布的《国家人工智能产业综合标准化体系建设指南(2024 版)》指出，到 2026 年，我国要“新制定国家标准和行业标准 50 项以上”、“参与制定国际标准 20 项以上，促进人工智能产业全球化发展”。

表 3　我国人工智能领域的技术标准

名称	发布机构	发布时间
《国家新一代人工智能标准体系建设指南》	国家标准化管理委员会等五部门	2020 年 7 月 27 日
《信息技术 人工智能 术语》（GB/T 41867-2022）	市场监督管理总局等两部门	2022 年 10 月 12 日
《信息化标准建设行动计划（2024-2027 年）》	网信办等三部门	2024 年 5 月 29 日
《国家人工智能产业综合标准化体系建设指南（2024 版）》	工信部等四部门	2024 年 6 月 5 日

（四）安全监管和评估

我国在多方布局人工智能技术发展的同时，同样重视对人工智能领域相关算法、技术的安全监管与评估，尤其是涉及国家安全、社会稳定以及个人权益等方面。《规划》同样要求应建立健全公开透明的人工智能监管体系，促进人工智能行业和企业自律。尤其是在生成式人工智能技术快速发展并被广泛应用的背景下，我国更是率先出台了对生成式人工智能服务管理的专门规定。目前，我国在人工智能安全监管与评估领域出台的主要法律政策包括：

表 4　我国生成式人工智能服务管理规定

名称	发布机构	发布时间
《关于加强互联网信息服务算法综合治理的指导意见》	网信办等九部门	2021 年 9 月 17 日
《互联网信息服务算法推荐管理规定》	网信办等四部门	2021 年 12 月 31 日
《互联网信息服务深度合成管理规定》	网信办等三部门	2022 年 11 月 25 日
《关于深入推进跨部门综合监管的指导意见》	国务院办公厅	2023 年 1 月 13 日
《生成式人工智能服务管理暂行办法》[4]	网信办等七部门	2023 年 7 月 10 日
《网络安全标准实践指南——生成式人工智能服务内容标识方法》	全国网络安全标准化技术委员会	2023 年 8 月 25 日
《生成式人工智能服务安全基本要求》	全国网络安全标准化技术委员会	2024 年 2 月 29 日
《信息安全技术 生成式人工智能数据标注安全规范》（征求意见稿）	全国网络安全标准化技术委员会	2024 年 4 月 3 日
《信息安全技术 生成式人工智能预训练和优化训练数据安全规范》	全国网络安全标准化技术委员会	2024 年 4 月 3 日

三、美国人工智能法律政策

（一）美国人工智能监管法律政策概述

美国目前尚未有联邦层面的人工智能监管立法，当前主要通过行政部门行政命令的形式，辅以各州自行立法加特定领域立法的方式对人工智能进行监管。其中拜登总统签署的《安全、可靠、可信地开发和使用人工智能行政令》明确了美国政府对待人工智能的政策法制框架。此外，白宫科技政策办公室在 2022 年 10 月发布《人工智能权利法案蓝图》，旨在保护公众在人工智能时代的权利，并确保自动化系统在设计、使用和部署过程中能够符合民主价值观，保护公民权利、自由和隐私。以上两份文件可以视为美国对人工智能监管的基本框架与路线图。美国人工智能相关法律政策主要包括：

表 5　美国人工智能相关法律政策

名称	法案简介	发布时间
《人工智能增长研究法案》	旨在加速美国经济和国家安全的人工智能研发，并缩小现有的资金缺口	2019 年 4 月

[4] 《生成式人工智能服务管理暂行办法》以“暂行”形式体现出短期立法的特征，说明亦将随技术和认识深入，有快速调整的可能。

《生成人工智能网络安全法案》	要求美国商务部和联邦贸易委员会（Federal Trade Commission）明确人工智能在美国应用的优势和障碍；调查其他国家的人工智能战略，并与美国进行比较；评估供应链风险以及如何解决这些风险。	2020 年 5 月
《2020 年国家人工智能倡议法案》	强调要进一步强化和协调国防、情报界和民用联邦机构之间的人工智能研发活动；设立国家人工智能倡议办公室，承担“监督和实施美国国家人工智能战略”等职责	2021 年 1 月
《人工智能权利法案蓝图》	设计、使用和部署自动化系统的五项原则，从而在人工智能时代保护美国公众	2022 年 10 月
《促进美国人工智能法》	要求特定的联邦机构采取措施促进人工智能的使用	2022 年 12 月
《安全、可靠、可信地开发和使用人工智能行政令》	为人工智能安全和保障制定了国内新标准，保护美国人的隐私，促进公平和公民权利，维护消费者和工人的利益，促进创新和竞争	2023 年 10 月
《编辑和深度伪造媒体内容来源保护和完整性法案》	要求为标记、验证和检测人工智能生成的内容制定新的联邦透明度准则，保护记者、演员和艺术家免受人工智能（对智力成果）的“剽窃”，并追究违法者滥用权利的责任	2024 年 7 月
科罗拉多州《SB205 法案》	专门用以识别、记录和减轻算法歧视风险	2024 年 6 月

（二）典型法律文件对人工智能风险问题的提出与回应

《安全、可靠、可信地开发和使用人工智能行政令》将人工智能定义为：一种基于机器的系统，可以针对一组人类定义的目标做出预测、推荐或决策，从而影响真实或虚拟环境。人工智能系统使用机器和人类输入来感知真实和虚拟环境，通过自动化方式分析这些感知，并使用模型推理来制定信息或行动的选项。

美国对人工智能的监管，强调安全保障、促进创新、公平与反歧视、隐私保护、可解释与透明度、保护消费者利益、保护本土就业岗位等基本原则，并关注人工智能对人类的替代与控制。在监管思路上，美国重视通过“红队测试”、标准化与技术评估、公众参与等方式来确保人工智能系统的安全。

《安全、可靠、可信地开发和使用人工智能行政令》提出美国要解决新兴知识产权问题，保护发明者和创造者的权益，通过立法和政策调整应对人工智能技术带来的知识产权挑战，并支持创新和公平竞争。美国专利商标局（USPTO）的相关指南和政策，明确了人工智能在发明过程中的角色，以及如何处理人工智能生成的发明，以确保人工智能系统的发明者能够获得专利保护。此外，《人工智

能权利法案蓝图》强调了防止大公司利用市场地位对小企业和初创公司造成不公平竞争，确保所有市场参与者都有公平的机会利用人工智能技术进行创新。

美国对人工智能系统可能涉及的人格权保护，主要从数据隐私的角度入手。但当前美国并没有一部联邦层面个人信息保护的立法，需要在部署人工智能时参考各州、各行业对个人信息保护的立法规定。此外，《人工智能权利法案蓝图》中，专门强调了透明性和可解释性，要求人工智能系统的决策过程应透明和可解释，让用户理解系统的工作原理和决策依据。《人工智能权利法案蓝图》还要求人们能够免于算法的歧视，要求防止基于种族、性别、年龄等的算法歧视，确保所有人平等享有机会和权利。

美国对透明度的细化要求主要通过标准文件落实，美国国家标准与技术研究院公开征求意见的（NIST）的 NIST AI 600-1 Generative AI Profile 文件是《安全、可靠、可信地开发和使用人工智能行政令》的配套标准，拟使用可解释的机器学习技术，使人工智能过程和结果更透明，便于理解决策是如何做出的。并应用透明度工具如数据表、数据营养标签和模型卡，记录解释和验证信息。在审查时，建议考虑透明度工件如影响评估、系统卡、模型卡和传统风险管理文档，作为组织决策的一部分。《安全、可靠、可信地开发和使用人工智能行政令》强调识别和标记人工智能生成的合成内容的重要性，并提出在内容中使用水印等技术来标记合成内容，并计划确定相关的标准、工具、方法和实践。

为了推动人工智能创新，白宫要求国家科学基金会（NSF）启动一个试点项目，实施国家人工智能研究资源（NAIRR），整合计算、数据、模型和训练资源，以支持人工智能相关的研究与开发。并且要求小企业管理局（SBA）优先分配区域创新集群计划资金，用于支持人工智能创新与商业化的规划活动。此外，还将评估现有计划的资格标准，以更好地支持小企业在采用人工智能方面的费用。为了吸引全球人工智能人才，美国政府简化签证申请处理时间，确保有足够的签证预约数量以吸引非公民在美国从事人工智能相关的工作、学习或研究。

关于人工智能伦理，《人工智能权利法案蓝图》特别强调了公平、公正、透明和负责的人工智能开发和使用的原则。主要关注点包括避免算法歧视、保护隐私、确保系统的透明度和提供人性化的选择。文件中指出，自动化系统不应基于种族、性别、宗教等进行不公正的区别对待。自动系统应在使用前进行测试，以

确保其没有算法歧视，并设计以确保广义上的公平。还提到人性化选择的必要性，用户应有权选择人工替代方案，并能够在自动系统出错时快速获得人性化的解决方案。

（三）监管机构设施和体现的分散性立法特点和分歧

美国对人工智能进行多头监管，其中白宫科技政策办公室在制定国家人工智能政策方面发挥领导作用，协调各联邦机构的人工智能活动和政策。国土安全部负责评估人工智能对国家安全的潜在威胁，并制定应对措施以确保安全。联邦通信委员会监管人工智能在通信领域的应用，特别是涉及无线通信和广播的人工智能技术。联邦贸易委员会负责保护消费者免受不公平或欺诈性商业行为的侵害，这包括监督人工智能在商业应用中的使用，确保其公平性和透明度，并对使用人工智能技术进行不公平或欺骗性行为的公司采取执法行动。商务部通过其下属的国家电信和信息管理局（NTIA）和其他部门，对人工智能技术的发展和使用进行监督。NIST 在人工智能技术标准的制定和风险管理框架的开发中起着关键作用。其任务是制定行业标准，确保人工智能技术的安全性和可靠性。

值得注意的是，除了在联邦和州层面存在法律政策进度异步的情况外，不同的州在人工智能的法律政策方面亦存在较大差异。典型事例是 2024 年 10 月，由于科技界的强烈反弹，加州州长纽森否决了备受关注的人工智能 AI 安全法案（SB1047）。其背后的科技与法律博弈的逻辑就在硅谷公司普遍认为：过于干预了仍处于发展阶段的 AI 技术，可能抑制创新。

总之，美国基于在发展方面处于优势地位，技术水平和研发能力较为领先的发展地位，对人工智能的监管法规要求维持和强化技术优势地位，同时强调安全性、透明度、公平性和非歧视，要求标识和认证人工智能生成内容，保护消费者隐私和数据安全。法规要求高风险人工智能系统进行定期审查和信息披露，多部门协作制定和执行监管政策，确保人工智能技术的负责任开发和使用，同时促进技术创新和保障国家安全。

四、欧盟人工智能法律政策

（一）欧盟的典型立法

欧盟立法的主要动因是为了应对人工智能技术带来的风险和挑战，同时促进技术创新和经济发展。2019 年 4 月，欧盟委员会出台《人工智能伦理准则》，确定了人工智能治理的尊重人类尊严、防止伤害、公平和可解释性四项伦理原则，并提出了实现这些原则的七项要求，即：尊重人类自主权、预防伤害、公平性和非歧视性、透明度和可解释性、隐私和数据治理、社会和环境福祉、问责机制。欧盟希望通过立法建立一个统一的法律框架，确保人工智能技术的应用符合欧盟的价值观和基本原则。2024 年 5 月 21 日，欧盟理事会正式批准了全球首部《人工智能法案》，该法案旨在协调人工智能监管规则，法案在经欧盟议会和欧盟理事会主席签署后，将于近日在欧盟官方公报上发布，并在公布 20 天后生效。欧盟《人工智能法案》的正式通过标志着全球人工智能领域监管迈入全新时代。

表 6　欧盟人工智能相关法律政策

名称	内容简介	发布时间
《欧盟机器人民事法律规则》	对基于人工智能控制的机器人，提出其使用的责任归属、伦理规则及对人类的伤害赔偿等监管原则，呼吁欧盟委员会评估人工智能的影响	2016 年 5 月
《人工智能协调计划》	提出了协调欧盟机构与成员国政府共同推进的 8 项具体任务：完善战略布局、加强多元投入、促进产研结合、培养专业人才、推动数据共享、建立监管框架、鼓励应用、加强国际合作	2018 年 12 月
《人工智能伦理准则》	确定了人工智能治理的四项伦理原则：尊重人类尊严、防止伤害、公平和可解释性	2019 年 4 月
《人工智能白皮书：通往卓越与信任的欧洲之路》	提出以监管和投资为导向，实现促进人工智能应用和解决应用风险的双重目标；人工智能监管框架应以风险为导向，基于风险评估对人工智能应用分级分类，采取不同的事前规制	2020 年 2 月
《人工智能法案》	根据人工智能的潜在风险和影响程度进行人工智能系统分级，并规定了人工智能参与者的义务，旨在建立一套统一的规范和监管框架，以确保人工智能技术的发展和应用能够遵循公平、透明和可信的原则	2024 年 5 月

主要有如下 6 个角色会受到《人工智能法案》的影响，分别是：Provider 提供者、Deployer 部署者、Distributor 分发者、Importer 进口者、Authorised representative 授权代表、Product manufacturer 产品制造商。具体说明如下：

Provider 提供者：开发人工智能系统或通用人工智能模型（或已开发人工智能系统或通用人工智能模型）并将其投放市场或以其自己的名称或商标投入使用的自然人或法人、公共当局、机构或其他机构，无论是付费还是免费；

Deployer 部署者：在其权力范围内使用人工智能系统的任何自然人或法人、公共当局、机构或其他机构，但用于个人非专业活动的人工智能系统除外；

Distributor 分发者：供应链中除提供者或进口者以外的任何将人工智能系统投放到联盟市场的自然人或法人；

Importer 进口者：位于或设立在联盟内的任何自然人或法人，将带有在联盟以外设立的自然人或法人的名称或商标的人工智能系统投放市场；

Authorised representative 授权代表：位于或设立在联盟内的任何自然人或法人，已收到并接受人工智能系统或通用人工智能模型提供者的书面授权，分别代表其履行和执行本法规规定的义务和程序。

Product manufacturer 产品制造商：将人工智能系统与其产品一起以自己的名称或商标投放市场或投入使用；

（二）法案对人工智能系统的监管要求

鉴于人工智能解决方案和应用范围广泛且不断变化，欧盟《人工智能法案》对人工智能的定义同样广泛：基本上，任何在欧盟部署的基于数据的驱动系统，无论其开发地点或数据来源如何，都将受其监管。而《人工智能法案》的独特之处在于其分级的体系设计以及基于相对风险的差异化义务设定。法案共将人工智能系统风险类型分为 4 类：一、不可接受的风险类型；二、高风险类型；三、有限风险或轻微风险类型和四、低风险类型。每一类风险类型的管理要求如下：

1、禁用不可接受的风险类型的人工智能系统

被认为对人类构成威胁的人工智能系统属于不可接受风险的类别。例如：社会评分系统、旨在操纵儿童或其他弱势群体的系统、实时远程生物特征识别的系统。以上这些系统是被禁用的。

2、重点监管高风险类型人工智能系统

对安全或基本权利产生负面影响的人工智能系统属于高风险。高风险可分为两个子类别：

（1）人工智能系统用于欧盟产品安全法规范围内的产品，包括玩具、航空、汽车、医疗设备和电梯等。

（2）需要在欧盟数据库中注册的八个领域的人工智能系统，包括：自然人的生物特征识别和分类；关键基础设施的管理和运营；教育和职业培训、就业、工人管理和自主创业；获得和享受基本私人服务、公共服务及福利；执法；移民、庇护和边境管制管理；协助法律解释和法律适用。

对于高风险系统，必须在上市前和整个生命周期内进行评估。

3、有限风险或轻微风险类型的人工智能系统，有较高自由度

有限风险类型的人工智能系统，是指使用者在应用系统时能够意识到在与人工智能互动，且使用者仍然能以自己的判断做出明智的决定。

该风险类型的人工智能系统在欧盟法案的监管框架下有较高的自由度，投放市场或投入使用前无需取得特殊的牌照、认证或履行繁杂的报告、监督、记录留存等义务。当前市场上常见的聊天机器人、文字和图片识别及生成软件、人工智能伴侣等大多属于此一风险类型。

4、低风险类型的人工智能系统，没有特殊的审查制度

未归类在不可接受的风险、高风险或有限风险类型的其他人工智能系统，都属于轻微风险类型。该风险类型的人工智能系统没有特殊的干预和审查制度，但提供者可自愿建立各行业的行为准则。

此外，《人工智能法案》对于高风险系统提出了合格评定（CA）的要求。合格评定的目标是验证高风险系统是否符合《人工智能法》中规定的七项要求，即风险管理、数据治理、技术文档、记录保存、透明度义务、人工监督以及准确性、稳健性和网络安全。除非另有规定，应在人工智能系统投入使用或进入市场之前满足所有这些要求。一旦系统投入使用，供应商还必须确保在系统的整个生命周期内持续合规。

（三）合规处罚与监管要求

针对有可能出现的违法行为，法案规定了较大的处罚力度。根据违法行为的

严重程度，法案设计了阶梯状的处罚条款：例如，违反禁止的人工智能实践规定，将被处以最高 3500 万欧元的行政罚款，如果违法者是企业，则最高罚款金额为其上一财年全球年营业额的 7%，以较高者为准。值得注意的是，在每一级的处罚阶梯上，法案都规定处罚实际金额将在固定金额和营业额比例金额中取其高者。按照法案“就高不就低”的规定，企业的违法成本异常高昂。

（四）实施展望

法案要求所有高风险人工智能系统必须进行基本权利影响评估，并在公共数据库中注册。同时，法案通过设立人工智能监管沙盒，鼓励创新技术的开发和测试，确保这些创新在安全和可控的环境中进行。法案还设立了多个监管机构，以确保法律的实施和遵守。

《人工智能法案》具有域外效力。它适用于任何在欧盟内部运营人工智能系统的公司以及位于欧盟以外的公司。

五、主要比较维度与框架映射

基于国别监管政策、法律体现的不同侧重和关注，本报告尝试以中国《生成式人工智能服务管理暂行办法》为基准，比较中美欧典型法律、政策文件在人工智能法律质量（立法和政策强度、弹性）方面的可能涉及或触发的 21 个维度，并投射到全国网络安全标准化技术委员会 2024 年 9 月发布的《人工智能安全治理框架》（“框架“）的 26 个风险控制项下。

表 7 中美欧人工智能典型法律政策文件比较

序号	比较项	中国	欧盟	美国	框架
	主要比较依据	生成式人工智能服务管理暂行办法	人工智能法	安全、可靠、可信地开发和使用人工智能行政令	
1	原则	发展和安全并重、促进创新和依法治理相结合	以人为本和支持创新	八项原则：人工智能必须安全可靠、促进负责任的创新、竞争和合作将使美国在人工智能领域处于领先地位、负责任地开发和使用人工智能需	包容审慎、确保安全；风险导向、敏捷治理；技管结合、协同应对；开放合作、共治共享

				要承诺支持美国工人、致力于促进公平和公民权利、保护公众（消费者）利益、保护隐私与公民自由、管理联邦政府使用的人工智能风险，提高监管、治理和负责任的使用的能力、引领全球社会、经济和技术进步	
2	定义	生成式人工智能技术，是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术	（人工智能系统）基于机器的系统，设计为以不同程度的自主性运行，在部署后可能表现出适应性，并且为了明确或隐含的目标，从其接收的输入中推断如何生成可影响物理或虚拟环境的输出，如预测、内容、建议或决定	基于机器的系统，其针对给定的一组人类定义的目标，做出预测、建议或影响真实或虚拟环境决策。人工智能系统使用基于机器和人类输入感知真实和虚拟环境；通过自动化分析将感知抽象为模型；并使用模型推理制定信息或行动选择；并进一步定义了模型和系统，以及生成式人工智能	
3	监管思路	包容审慎和分类分级监管	基于风险分级和风险控制	基于两用性和风险管理，特别是强制红队测试等；行政令始终贯穿风险优先方向，并要求通过制定国家安全备忘录等方式协调评估人工智能的双重用途	风险分类
4	价值观	核心价值观	欧盟宪章的价值观	国家安全、公共健康和安全利益	
5	监管机构	网信、工信、公安、科技、发改等	人工智能办公室、成员国代表组成的欧洲人工智能委员会、科学小组、咨询论坛；成	白宫人工智能委员会，协调既有的主要联邦政府部门的人工智能政策	

			员国至少一个通知机关和至少一个市场监督管理机关		
6	监管沙盒/真实世界测试	已有地方和若干行业开展的监管沙盒先行先试的实践	要求成员国建立至少一个国家级人工智能监管沙盒，规定了监管目的、原则、模式，以及透明度、豁免等条件；规定进行真实世界测试的条件和数据保护要求等（计划及其批准、时限、知情同意、检查、数据保护等）		作为“预监管”或先行先试的措施
编号	风险项	生成式人工智能服务管理暂行办法	人工智能法	安全、可靠、可信地开发和使用人工智能行政令	人工智能安全治理框架
1	内容安全（禁止的活动）	国家、社会公共利益，其他合法权益（十不准）	通过定义八项禁止类人工智能实践进行限制（其他非法、虚假、歧视）	基于行政令原则范围的国家安全、偏见、歧视等；对外国恶意网络行为者使用美国基础设施即服务（IAAS）产品的禁令（如立法程序中的 ENFORCE 法案）	信息内容安全风险 4.2.1 （a）
2	反歧视	应在算法设计、训练数据选择、模型生成和优化、提供服务等过程中，采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视	高风险人工智能系统应审查可能存在的偏差，这些偏差可能……，或导致欧盟法律所禁止的歧视	围绕算法歧视解决涵盖侵犯自由和种族、社群等在社会福利、用工、租赁和住房、医疗、教育等领域歧视问题；已制定《关于打击自动化系统中的歧视和偏见的执法工作的联合声明》；科罗拉多州颁布 SB205 法案专门用以识别、记录和减轻算法歧视风险	偏见、歧视风险 4.1.1 （b）

3	知识产权	应尊重知识产权、商业道德，保守商业秘密	（1）不影响欧盟和国内法的尊重和保护知识产权、商业秘密规定；（2）更新欧盟版权法的政策	（1）发布指南以解决人工智能的专利发明人、发明权问题；（2）提供解决作品的利用和保护的版权法建议	被窃取、篡改的风险 4.1.1（b）；训练数据含不当内容、被“投毒”风险 4.1.2（b）
4	促进竞争/反不正当竞争	不得利用算法、数据、平台等优势，实施垄断和不正当竞争行为	提供者不应通过采用低于欧盟规定的版权标准在欧盟市场上获得竞争优势；数字市场法等亦适用	通过 CHIPS 法案促进竞争，支持初创/小企业创新和商业化；奖金、贷款和基金等	诱发传统经济社会安全风险 4.2.2（b）
5	人格权	不得侵害他人肖像权、名誉权、荣誉权、隐私权等	明确欧盟宪章下的基本权利；GDPR 的个人数据保护要求继续适用；高风险类的人工智能系统提供者在处理特殊类别的个人数据，如个人数据的重复使用方面应受到技术限制，并采取先进水平的安全和隐私保护措施；此外，《值得信赖的人工智能的伦理准则》的隐私和数据治理准则亦适用	评估人工智能在刑事法律中的应用（包括预测犯罪），包括保障措施和使用限制；分析和了解、调查以法律形式剥夺公民权利的人工智能案件	挑战传统社会秩序的风险 4.2.4（a）（b）
6	个人信息	适用《个人信息保护法》（包括提供者应当依法及时受理和处理个人关于查阅、复制、更正、补充、删除其个人信息等的	GDPR 的个人数据保护要求继续适用，包括说明收集数据的初始目的等；对特殊个人数据的处理（如情感识别）至于禁止类人工智能系统；高风险人工智能系统提供者对特殊	（1）通过 OMB 牵头评估和对人工智能收集和使用个人信息采取措施；（2）支持相关隐私增强技术的设计、开发和部署（如隐私增强技术）；（3）将安全、隐私和安全标准纳入软件开发	违规收集使用数据风险 4.1.2（a）；数据泄露风险 4.1.2（c）（d）；不当使用引发信息泄露风险

		请求等）	个人数据的额外要求	生命周期	4.2.1（b）
7	透明度	应基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度（非强制性约束，对透明度无具体细化）	明确了统一的透明度规则。规定高风险人工智能系统的设计和开发应确保其操作具有足够的透明度……；增加特定人工智能系统的提供者和部署者的透明度义务	强调或澄清与人工智能模型透明度相关的要求和期望，以及受监管实体应解释其使用人工智能模型的能力；在包括能源部的人工智能模型评估工具和人工智能试验平台中配备具有透明度和可复制的工具和技术测试	可解释性的风险 4.1.1（a）
8	数据质量	应提高生成内容的准确性和可靠性（非强制性约束）	明确用于训练、验证和测试的高质量数据集需要实施适当的数据治理和管理实践；对高风险人工智能系统，应使用符合质量标准的训练、验证和测试数据集		偏见、歧视风险 4.1.1（b）； 鲁棒性弱风险 4.1.1（b）； 输出不可靠风险 4.1.1（a）（b）
9	国际规则	鼓励开展国际交流与合作，参与生成式人工智能相关国际规则制定（如《全球人工智能治理倡议》《布莱切利宣言》）	（1）强调包括数字贸易在内的国际贸易承诺；（2）考虑与从事人工智能的国际组织的成果对齐和可接受性，要求立法和政策应考虑国际进展；（3）推动技术和标准上的国际合作	推动在其盟友和合作伙伴的国际间合作和建立强有力的国际框架；发布全球人工智能发展手册；制定全球人工智能研究议程等	供应链安全风险 4.1.3（d）； 加剧社会歧视偏见、扩大智能鸿沟的风险 4.2.4（a）； 挑战传统社会秩序的风险 4.2.4（a）（b）
10		（1）使用具有合法来源的数据和基础模型；（2）涉及知识产	基于形式符合和涵盖上市前和上市后过程管理，高风险人工智能系统的提供者应履	（1）两用基础模型的开发者应持续向联邦政府提供相关信息、报告或记录，特别是应对复杂威	模型算法安全风险 4.1.1； 数据安全风险

	提供者/开发者一般义务	权的，不得侵害他人依法享有的知识产权；（3）涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形；（4）采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性；（5）依法承担网络信息内容生产者责任，履行网络信息安全义务；（6）对使用者的输入信息和使用记录应当依法履行保护义务；（7）应当在其服务过程中，提供安全、稳定、持续的服务，保障用户正常使用	行相应的风险管理、质量管理体系、合格性评估、登记、报告、纠正等义务	胁而采取的物理和网络安全保护措施、在相关人工智能红队测试中的性能表现和加强整体模型安全性的缓解措施、报告潜在大规模计算集群收购、开发或权属情况等；（2）境外恶意网络行为者使用美国基础设施即服务（IaaS）产品风险报告和禁止类义务等	4.1.2；系统安全风险 4.1.3；网络域风险 4.2.1；现实域风险 4.2.2；加剧“信息茧房”效应风险 4.2.3（b）；加剧社会歧视偏见、扩大智能鸿沟的风险 4.2.4（a）；挑战传统社会秩序的风险 4.2.4（a）（b）
11	数据标注	制定清晰、具体、可操作的标注规则；开展数据标注质量评估，抽样核验标注内容的准确性；对标注人	在技术文件中提供数据要求，说明训练方法和技术以及所使用的训练数据集，包括……标注程序		训练数据标注不规范风险 4.1.2（e）

		员进行必要培训			
12	使用者/部署者义务	（1）第四条的一般性义务（包括价值观、反歧视、透明度、数据质量等）；（2）使用者发现生成式人工智能服务不符合法律、行政法规和本办法规定的，有权向有关主管部门投诉、举报	高风险人工智能系统部署者的义务包括：（1）基于系统说明的正当使用；（2）人工监督；（3）向提供者报告高风险和暂停使用义务；（4）日志留存；（5）特殊高风险系统的知情同意、报告等义务；（6）基本权利影响评估等	与提供者等统一作为受监管实体，但具体和差异化的义务尚待在具体的生效法案中明确，例如在州法SB205中，已经区分提供者与部署者义务	供应链安全风险4.1.3（d）；网络域风险 4.2.2；现实域风险 4.2.2
13	防沉迷	应采取措施防范未成年人用户过度依赖或者沉迷	（空）	（空）	
14	标识	提供者应当按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识	生成合成音频、图像、视频或文本内容的人工智能系统……提供者应确保人工智能系统的输出以机器可读的格式进行标注，并且可检测其系人为生成或操纵	制定标注工具和实践指南，用以验证合成内容和内容检测，减少合成内容风险	混淆事实、误导用户、绕过鉴权的风险 4.2.1（a）
15	违法处置义务	（1）提供者发现违法内容的，应当及时采取停止生成、停止传输、消除等处置措施，采取模型优化训练等措施进行整改，并向有关主管部门报告；（2）	（1）高风险人工智能系统存在……风险，且提供者意识到该风险时，则应立即告知系统部署者，如适用……通知其提供高风险人工智能系统所在成员国的市场监督管理机关；（2）高风险人工智能	行政令主要旨在进行评估和指引，对违法行为的法律责任作为执法和司法机构的素养提升范围，如包括通过改进和增加对联邦执法官员、联邦检察官的培训，以调查和起诉与人工智能相关案件	

		发现使用者利用生成式人工智能服务从事违法活动的，应当依法依约采取警示、限制功能、暂停或者终止向其提供服务等处置措施，保存有关记录，并向有关主管部门报告	系统提供者应向发生事故的成员国市场监督管理机关报告任何严重事件；（3）具有系统风险的通用人工智能模型的提供者应……跟踪、记录并及时向人工智能办公室报告，并酌情向国家主管机关报告严重事件的相关信息以及为解决这些问题可能采取的纠正措施；（4）以及更为严格的召回等		
16	算法备案/登记/注册	提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应当……按照《互联网信息服务算法推荐管理规定》备案	高风险人工智能系统投放市场或提供服务之前，应在欧盟数据库中登记		基于人工智能（风险）分类分级的备案、登记
17	可解释性	无明确提法（在透明度和协助义务中，及《个人信息保护法》中有涉及）	（1）高风险人工智能系统的设计和开发应确保其操作具有足够的透明度，编制和更新的技术文件中包括对系统结构、输出等的可解释性；支持部署者能够解释系统的输出并加以适当使用（如系统提供与解释其产出相关的信息的技术能力和特点、解释日	与人工智能模型透明度相关	可解释性差的风险4.1.1（a）； 输出不可靠风险4.1.1（a）（b）

			志等）；（2）受到影响的个人有权要求部署者就人工智能系统在决策程序中的作用和所做决定的主要内容做出明确而有意义的解释		
18	伦理	适用《科技伦理审查办法（试行）》（如建立科技伦理（审查）委员会，建立健全全流程科技伦理监管机制和审查质量控制、监督评价机制，在开展科技活动前进行科技伦理风险评估或审查	适用《欧洲可信人工智能伦理准则》明确的人类主体和监督；技术稳健性和安全性；隐私和数据治理；透明度；多样性、非歧视和公平；社会和环境福祉以及问责制的伦理原则；真实世界测试高风险人工智能系统时应增加伦理审查	明确通过伦理观念和人员的引入，以管理联邦政府使用人工智能的风险，确保监管、治理和支持负责任的使用人工智能	伦理域风险 4.2.4
19	公共机构/关键基础设施（CI）运营者使用人工智能		在禁止类、高风险人工智能系统中对涉及公共利益（如公共机关为其自身使用）数据和应用作出考虑，并进行了必要约束（如要求开展基本权利影响评估）	（1）评估 CI 使用人工智能的风险并强制报告，同时利用人工智能改善 CI 的安全，如漏洞发现和修复能力；（2）联邦政府建立机构首席人工智能官等角色等等	网络域风险 4.2.1；现实域风险 4.2.2；伦理域风险 4.2.4
20	对模型的特殊考虑		将符合一定计算性能或其他专门评估的通用人工智能模型类型化为具有系统性风险的通用人工智能模型，并规定了专门义务，包括持续更新的技术文件、模型评估（含	明确提供者应对双重用途基础模型（权重）开展风险评估，和消除风险的保障机制；制定包括开发能源部的人工智能模型评估工具和人工智能试验平台的计划；为模型设计开发模型	模型算法安全风险 4.1.1；模型复用的缺陷传导风险 4.2.1（a）（b）

			对抗测试）、网络安全保护、行为准则等	护栏	
21	特殊领域的风险关注		特别关注公共场所为执法目的的生物识别风险	关注在特殊领域的应用风险，如化学、生物安全、放射性、核等；要求在卫生（包括上市前评估和上市后监督的保障政策）、交通、教育等若干主要部门建立人工智能应用的战略或指南类文件；打击和阻止不必要的机器人（自动拨号）电话等保护消费者权益；评估人工智能应用对劳动力市场的影响	用于违法犯罪活动的风险 4.2.2（a）（b）； 两用物项和技术滥用风险 4.2.2（a）（b）； 用于开展认知战的风险 4.2.3（a）（b）（c）； 挑战传统社会秩序的风险 4.2.4（a）（b）； 未来脱离控制的风险 4.2.4（b）

六、结论和展望

本报告主要收集和整理了三个主要立法区域截至 2024 年的部分主要监管思路、维度。需要明确的是，当前的人工智能技术仍在快速发展中，这一快速演化和不确定性构成了对政策法律监管的外部约束。从对主要国家的立法趋势观察，未来通过联合国等多边、双边机制，各国的政策法律在一些基础原则、国际规则上将可能产生某些普遍价值的一致性，例如联合国大会通过《抓住安全、可靠和值得信赖的人工智能系统带来的机遇，促进可持续发展》等法律文件，正在努力驱动和寻求各国人工智能的共识与协同；同时各国技术、产业发展的差异又将不断分化和催生出新的、阶段性的监管需求，这又导致了一定时期内，各国监管的重点和分歧，甚至博弈加剧和技术、监管的对抗性。我们也将保持持续关注更新。

CSA GCR cloud security alliance GREATER CHINA REGION ®

Cloud Security Alliance Greater China Region

扫码获取更多报告

云安全联盟大中华区
官网：https://c-csa.cn 电话：0755-86548359
地址：上海市浦东新区东育路255弄5号3FB30（前滩国际经济组织集聚区）/
广东省深圳市南山区粤海街道北京航空航天大厦1号楼905