



2025 年度全球人工智能治理报告

目录

执行摘要	1
一、跨国治理协议与机构框架：从共识构建走向实施协作	2
1. 主要国际组织的框架性工作	2
2. 关键多边治理协议与倡议	3
二、地区层面：多边集团的路径分化与主张协同	3
1. 欧盟：从立法到生态的体系化建设	4
2. 金砖国家 (BRICS)：双轨推进产业合作与规则塑造	4
3. 东南亚国家联盟 (ASEAN, 简称：东盟)：深化区域协作与中国务实合作	4
4. 中东地区：资本驱动的基础设施与枢纽竞赛	5
5. 拉丁美洲和加勒比地区：明确发展路径与寻求区域协作	5
6. 非洲联盟 (African Union, 简称：非盟)：聚焦本土转型的战略实施	6
三、国家层面：主权国家的战略竞争与特色实践	6
1. 中国：“发展与安全”并重的系统化布局	6
2. 美国：“内松外紧”的竞争性结盟战略	6
3. 其他国家的特色化路径	7
四、核心发展趋势总结	7
附：2025 年全球人工智能治理大事记	9
1 月	9
1. 美国发布第 14179 号行政令（消除美国人工智能领导力障碍）	9
2. 韩国颁布《人工智能发展及可信基础构建基本法》	9
3. 国际人工智能安全报告发布	9
2 月	10
4. OECD 发布 AI 时代数据获取与共享政策建议	10
5. 《ALECSO 人工智能伦理宪章》起草委员会在 ALECSO 总部召开会议	11
6. OECD 推出全球框架，监测 G7 广岛 AI 行为准则实施情况	11
7. 巴黎 AI 行动峰会发布全球治理宣言	12
3 月	12
8. 中国出台《人工智能生成合成内容标识办法》	12
9. 西班牙批准《人工智能法草案》并启动监管沙盒	13
4 月	13
10. 欧盟发布人工智能大陆行动计划	13
11. 非洲联盟以本土化 AI 战略赋能非洲发展与转型	14
12. IEEE 智能家居多模态 AI 标准立项	15
13. 沙特阿拉伯发布《全球人工智能枢纽法案》	15
14. 中国出台全球首部气象 AI 专项法规	15
15. 中央政治局就人工智能发展与监管进行集体学习 习近平总书记强调推动人工智能健康有序发展	16
16. 欧盟发布生成式 AI 研究使用指南，为科研人员提供行动方向	16
5 月	16
17. 美国废止“AI 扩散规则”并升级芯片管制	16
18. 金砖国家人工智能高级别论坛在巴西利亚举行，《金砖国家人工智能展望报告（2025）》发布	17

19. 日本首部人工智能专项法律出台	17
6月	17
20. 七国集团领导人发布关于人工智能促进繁荣的声明	17
21. 第三届联合国教科文组织全球人工智能伦理论坛在泰国曼谷举行	18
7月	18
22. 金砖国家元首签署“人工智能全球治理宣言”	18
23. 中国在2025世界人工智能大会推出人工智能治理与安全系列倡议	18
24. 欧盟发布《通用人工智能行为准则》	19
25. 欧盟委员会正式发布《通用人工智能模型提供商指南》	19
26. 国际电信联盟“AI向善全球峰会”推动国际治理新倡议	20
8月	20
27. 欧盟《人工智能法案》部分条款正式生效	20
28. 中国国务院印发《关于深入实施“人工智能+”行动的意见》	21
29. 联合国拉丁美洲及加勒比经济委员会发布《加强加勒比地区人工智能准备度》 报告	21
30. 联合国设立“AI独立国际科学小组”与全球对话机制	22
9月	22
31. ISO发布全球首个AI可解释性国际标准	22
32. EDPB发布首份《数字服务法》(DSA)与《通用数据保护条例》(GDPR)协同 适用指南	22
33. 北京发布《从人工智能到超级智能北京宣言》	23
34. 中国发布《人工智能安全治理框架》2.0版	23
35. 2025中国—东盟人工智能部长圆桌会议在广西南宁召开	24
36. 美英签署《科技繁荣协议》	25
37. 美国加州通过《前沿人工智能透明法案》	25
10月	25
38. 欧盟委员会发布《应用人工智能战略》	25
39. 欧盟委员会发布《欧洲科学人工智能战略》	26
40. 中国国家互联网信息办公室发布《政务领域人工智能大模型部署应用指引》	26
41. 美日签署《科技繁荣协议》	27
11月	27
42. 欧盟启动AI生成内容标记准则制定	27
43. 欧盟《数字综合法案》提案引发争议	27
44. 美国特朗普政府启动“创世纪计划”	28
45. 韩国政府发布《AI领域规制合理化路线图》	28
12月	28
46. 越南通过了《人工智能法》	28
47. 美国特朗普政府签署AI监管行政令	29

执行摘要

2025 年，全球人工智能治理格局从原则探讨和框架设计，全面进入规则实施、战略布局与联盟竞争的“实施元年”。本年度的核心特征是“规则落地”与“路径分化”，一个多中心、碎片化的治理新秩序正在加速形成。

国际层面，治理呈现出“机构夯实框架，多边凝聚共识”的双轨特点。主要国际组织（如联合国、OECD、ISO）致力于将原则转化为可操作的工具与标准。同时，关键多边协议密集出台，例如 G7 推动其《广岛 AI 行为准则》监测框架落地，巴黎峰会 62 国签署包容性宣言，以及金砖国家发布强调公平发展的《人工智能全球治理宣言》，共同勾勒出国际社会在协作与博弈中寻求平衡的复杂图景。

地区层面，主要多边集团基于自身利益与价值观，选择了差异显著的发展路径。欧盟通过《人工智能法案》生效及配套战略，确立了“立法先行、生态构建”的监管范式。金砖国家采取“产业合作”与“规则塑造”双轨并进，倡导建立更包容的全球框架。东盟、非洲联盟、中东及拉美地区则分别聚焦于务实合作、本土化转型、资本驱动的枢纽竞争和区域能力建设，展现出全球治理实践的多样性。

国家层面，主要行为体的战略竞争白热化。美国转向“国内松绑、对外结盟”策略，通过《科技繁荣协议》构建技术联盟并强化出口管制。中国坚持“发展与安全”并重，系统性推进“人工智能+”应用与敏捷治理。韩国、日本、沙特等国则通过特色立法与制度创新，成为塑造治理多样性的关键“中间力量”。

2025 年全球 AI 治理呈现五大趋势：规则强制化、格局阵营化、竞争系统化、路径多样化和工具技术化。本报告通过详实的月度大事记（均附官方来源链接）提供了全景式记录与分析，旨在成为政策制定者、行业领袖和研究人士理解未来 AI 治理走向的参考。

2025年，全球人工智能治理图景的复杂程度骤然提升。焦点从技术“能做什么”，转向规则“由谁而定”以及红利“如何分配”。以欧盟《人工智能法案》生效等一系列标志性事件为节点，国际社会正式步入了一个规则密集落地、战略加速对冲的新时期，新的秩序在博弈中萌发。本报告旨在系统梳理这一年关键进展，首先审视国际层面的协议与机构博弈，继而剖析主要地区集团的差异化路径选择，最后聚焦关键国家的核心战略竞合。通过这三个维度的透视，揭示表面动态之下的深层逻辑与未来趋势。

一、跨国治理协议与机构框架：从共识构建走向实施协作

2025年，全球人工智能治理在国际层面呈现出“机构夯实框架，多边凝聚共识”的双轨特点。主要国际组织致力于将治理原则转化为可操作的规则与工具，重要的多边峰会催生了具有广泛代表性的治理倡议，共同应对AI带来的全球性挑战。

1. 主要国际组织的框架性工作

核心国际组织在制定标准、提供公共产品和搭建对话平台方面发挥了基础性作用：

- **联合国 (UN)**: 通过决议设立“人工智能独立国际科学小组”及“全球人工智能治理对话”常设论坛，旨在为全球治理提供权威科学评估与包容性政府间对话平台，强化以联合国为核心的多边治理主渠道。

- **国际电信联盟 (ITU)**: 在日内瓦举办“人工智能向善全球峰会”，发起“人工智能技能联盟”、“全球粮食系统人工智能倡议”等务实工具，重点帮助发展中国家培养人才、制定本地标准，以应对数字鸿沟挑战。

- **国际标准化组织 (ISO)**: 与 IEC 联合发布全球首个 AI 可解释

性国际标准《ISO/IEC TS 6254:2025》，为机器学习模型的透明度和可理解性提供了统一的技术框架与评估方法，标志着技术治理进入标准化阶段。

● **经济合作与发展组织（OECD）**：发布了《AI时代下增强数据获取和共享的政策建议》，推出了用于监测《G7广岛人工智能行为准则》实施情况的企业报告框架，将自愿性准则转化为可测量、可比较的合规实践。

2. 关键多边治理协议与倡议

2025年，由不同国家集团主导的多边倡议密集出台，成为塑造全球AI治理规则与联盟格局的核心力量。

（1）机制化治理框架的推进：G7与OECD在2月合作推出了监测《广岛AI行为准则》实施情况的全球报告框架，为企业提供了透明化、可比较的风险管理报告标准。6月，G7进一步发布《人工智能促进繁荣声明》，宣布建立“七国集团人工智能网络(GAIN)”，旨在系统性加速AI在公共部门及中小企业的应用。

（2）发展路径与治理理念的宣言：2月，巴黎AI行动峰会上62国签署的《包容与可持续发展宣言》，与7月金砖国家领导人发布的《关于人工智能全球治理的宣言》，分别代表了国际社会对“AI普惠发展”和“公平治理框架”的两大集体诉求。同期，在2025世界人工智能大会上发布的《人工智能全球治理行动计划》，系统性地提出了涵盖发展、安全、基建与合作的13条行动举措。

这些协议共同表明，全球AI治理正在形成“发达国家集团推动标准互认”、“新兴大国倡导规则重构”、“多边平台寻求最大公约数”并行的复杂局面。

二、地区层面：多边集团的路径分化与主张协同

由多国构成的地区性集团，基于共同价值观或利益，形成了差异

鲜明的治理路径与集体主张。

1. 欧盟：从立法到生态的体系化建设

欧盟已全面进入规则实施与战略支撑的新阶段。其核心进展一方面表现为**法规生效与细化**：全球首部综合性《人工智能法案》关于通用人工智能（GPAI）的关键条款于8月2日正式生效。同时，配套的《通用人工智能模型提供商指南》和自愿性的《通用人工智能行为准则》发布，构成了“法律-解释-实践”三层落地的监管体系。另一方面表现为**战略生态构建**：相继推出《人工智能大陆行动计划》、《应用人工智能战略》与《欧洲科学人工智能战略》，形成从基础设施、产业应用到前沿科研的完整战略拼图，旨在系统性提升欧盟的技术主权与竞争力。

2. 金砖国家 (BRICS)：双轨推进产业合作与规则塑造

2025年，金砖国家通过“产业合作”与“规则塑造”双轨并进的方式，深化了人工智能领域的共同立场。在产业合作层面，5月在巴西利亚举办了人工智能高级别论坛，发布了《金砖国家人工智能展望报告(2025)》等务实成果。在规则塑造层面，7月领导人会晤通过《人工智能全球治理宣言》，明确了核心主张：建立以联合国为中心的包容性全球治理框架，强调保障发展中国家的数字主权、发展权及对技术数据的公平获取，并推动AI服务于可持续发展。这标志着金砖国家正积极为全球南方争取人工智能时代的更大话语权。

3. 东南亚国家联盟 (ASEAN, 简称：东盟)：深化区域协作与中国务实合作

2025年，东盟在《东盟人工智能治理与伦理指南》(2024年发布)的基础上，进一步深化人工智能治理的区域协作与实践。与中国共同

建立了机制化的高层合作平台。9月，双方在广西南宁成功召开了首届“中国—东盟人工智能部长圆桌会议”。会议宣布启动建设“中国—东盟国家人工智能应用合作中心”，并提出了开展中国—东盟“人工智能+”行动，旨在从政策对话、产业对接与人才培养等方面推动务实项目落地，标志着双方 AI 合作进入了实体化、项目化的新阶段。

4. 中东地区：资本驱动的基础设施与枢纽竞赛

2025年，以沙特、阿联酋、卡塔尔为代表的海湾国家，通过巨额资本投入与制度创新，在全球人工智能基础设施与枢纽地位的竞争中表现突出。其中，沙特于4月发布《全球人工智能枢纽法案》草案，核心是创设法律地位特殊的“数据使馆”，旨在以制度优势吸引全球数据与投资，并同步推进先进数据中心建设。阿联酋继续推进其庞大的阿布扎比 AI 园区计划，并从美国获得高端 AI 芯片供应保障；卡塔尔通过主权基金专注基础设施投资，并发布了体现阿拉伯文化特性的 AI 模型。与此同时，阿拉伯国家也在协同构建本土治理框架，阿拉伯教科文组织(ALECSO)于2月召开了《ALECSO 人工智能伦理宪章》起草会议，致力于制定符合地区文化价值的伦理规范。

5. 拉丁美洲和加勒比地区：明确发展路径与寻求区域协作

拉丁美洲和加勒比地区在 2025 年的主要进展体现为通过国际报告明确发展路径，整体仍处于早期阶段。8月，联合国拉丁美洲及加勒比经济委员会发布了针对加勒比地区的政策简报，系统分析了 AI 应用的潜力(如灾害预警、智慧农业)与主要风险(如基础设施薄弱、人才短缺、法规不健全)，并提出了加强区域协作、提升治理能力的政策建议。区域内部发展不平衡，智利、巴西等国处于领先地位，但整体在数字基建、人才和数据利用能力上仍面临共同挑战。年内多次会议都强调需要建立符合区域特点的伦理框架和监管模式，并倡导设

立“区域 AI 观察站”等协作机制。

6. 非洲联盟 (African Union, 简称: 非盟): 聚焦本土转型的战略实施

积极推动《大陆人工智能战略》进入实施阶段, 强调 AI 必须服务于非洲本土的社会经济转型议程 (如农业、医疗), 并致力于开发体现非洲语言文化多样性的解决方案, 以维护数字主权。

三、国家层面: 主权国家的战略竞争与特色实践

主要主权国家依据自身战略定位, 推出了更具针对性、竞争性的国家战略与监管规则。

1. 中国: “发展与安全” 并重的系统化布局

中国在推进 AI 全方位应用的同时, 加速构建兼顾发展与安全的治理体系。

(1) 战略升级: 国务院印发《关于深入实施“人工智能+”行动的意见》, 设定了到 2035 年的中长期目标, AI 发展被纳入国家顶层系统工程。

(2) 监管细化与标准引领: 年内密集出台《人工智能生成合成内容标识办法》、《人工智能气象应用服务办法》等多个垂直领域法规。同时, 在智能家居等领域牵头制定 IEEE 国际标准, 实现从技术应用到规则输出的跨越。

(3) 积极参与全球治理: 2025 世界人工智能大会发布《人工智能全球治理行动计划》与《中国人工智能安全承诺框架》2.0 版, 主动贡献治理方案。

2. 美国: “内松外紧” 的竞争性结盟战略

美国采取了以维护全球技术领导力为核心、国内与国际联动的策略。

(1) 对内松绑与聚焦：撤销上届政府监管框架，签署《消除美国人工智能领导地位的障碍》行政令，并通过新行政令限制各州监管，旨在为私营部门“松绑”。同时，启动国家级“创世纪计划”(Genesis)，整合国家资源推动 AI 与科学发现深度融合。

(2) 对外结盟与遏制：与英国、日本先后签署《科技繁荣协议》，核心是构建以美国技术标准与安全规范为主导的盟友圈。在供应链上，升级对华高端 AI 芯片出口管制，构建“小院高墙”技术壁垒。

3. 其他国家的特色化路径

(1) 韩国：颁布整合国家战略、产业促进与风险监管的《人工智能发展及可信基础构建基本法》(2026 年生效)，并配套推出《AI 领域规制合理化路线图》，以“立法先行、系统备战”模式寻求突破。

(2) 沙特阿拉伯：发布《全球人工智能枢纽法案》草案，创新性提出设立法律地位特殊的“数据使馆”，旨在通过极致的制度创新吸引全球数据与资本，竞争区域 AI 枢纽地位。

(3) 日本：出台首部 AI 专项法律，设立首相牵头的“AI 战略本部”，并明确企业配合政府调查 AI 滥用的义务，体现“安全保障下的创新”思路。

(4) 新加坡：在金融等关键领域推出全球最前沿、最具体的《AI 风险管理指南》咨询文件，展现“垂直精耕”的治理风格。

四、核心发展趋势总结

综合而言，2025 年全球 AI 治理呈现以下趋势：

1. 规则强制化：全球治理已确定性从“原则讨论”步入“规则落地”时代，具有法律约束力的合规要求成为新常态。

2. 格局阵营化：治理理念深度分裂，形成“美国技术联盟”、“欧盟规则范式”与“中国发展道路”三足鼎立，同时与“全球南方发展

权诉求”交织互动的复杂格局，统一框架难觅。

3. 竞争系统化：竞争超越单纯技术研发，扩展至治理规则、标准联盟、数据基础设施、供应链安全与人才吸引的全体系竞争。

4. 路径多样化：除大国外，韩国、沙特等“中间力量”通过特色立法与制度创新，成为塑造治理实践多样性的关键变量。

5. 工具技术化：治理机制对技术工具的依赖空前加深。水印、可解释性框架等不再仅是可选功能，而是作为合规要件被写入法规与国际标准，推动监管与产业技术栈深度融合，显著提升了治理的“技术含量”与合规门槛。

2025 年是全球 AI 治理的“实施元年”与“分化元年”。一个基于不同价值观、由多行为体共同塑造的碎片化、竞争性治理格局已稳固成形。未来，在主导权争夺与发展权诉求的持续张力下，全球合作将更趋于在具体议题和功能性领域寻求有限共识。

附：2025 年全球人工智能治理大事记

1 月

1. 美国发布第 14179 号行政令(消除美国人工智能领导力障碍)

摘要：2025 年 1 月 20 日，美国特朗普政府撤销拜登-哈里斯政府 2023 年 10 月 30 日发布的第 14110 号 AI 行政命令《安全、可靠和可信赖的人工智能开发与使用》¹。特朗普政府认为，拜登政府的行政命令通过施加繁重的政府要求，限制了私营部门在 AI 领域的开发和部署，从而阻碍了创新。

2025 年 1 月 23 日，美国总统特朗普签署第 14179 号行政命令《消除美国人工智能领导地位的障碍》²，指示制定一项 AI 行动计划，以提升美国在 AI 领域的优势。

2. 韩国颁布《人工智能发展及可信基础构建基本法》

摘要：2025 年 1 月 21 日，韩国《人工智能发展及可信基础构建基本法》(简称《人工智能基础法》)³正式颁布，设定 1 年过渡期，将于 2026 年 1 月 22 日起全面生效。然而，该法案的核心设计也引发了显著争议。批评观点认为：该法案总体上正确把握了策略与产业政策方向，但其采用生硬的监管条款误判了人工智能风险，反而削弱了这些策略的效力。过于宽泛的定义、僵化的研发要求及中小企业优先规则，可能会扼杀韩国保持全球竞争力所需的规模与灵活性。诸如标签要求、计算阈值及流程繁琐的报告等生硬手段，将浪费资源却无法实现真正的责任落实⁴。

3. 国际人工智能安全报告发布

摘要：2025 年 1 月 29 日，人工智能行动峰会发布《国际人工智

¹ The White House. <https://www.whitehouse.gov/presidential-actions/2025/01/initial-rescissions-of-harmful-executive-orders-and-actions/>.

² The White House. <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>.

³ CSET. <https://cset.georgetown.edu/publication/south-korea-ai-law-2025/>.

⁴ ITIF. <https://itif.org/publications/2025/09/29/one-law-sets-south-koreas-ai-policy-one-weak-link-could-break-it/>.

能安全报告》(*International AI Safety Report*)⁵，这是首份国际人工智能安全报告，来自 30 个国家提名的国际专家顾问小组、经济合作与发展组织 (OECD)、欧盟和联合国的 96 位人工智能专家组共同参与完成该报告撰写，包括中国科学院自动化研究所曾毅研究员等。报告重点关注通用人工智能，总结了三个核心问题的科学证据：通用型人工智能能做什么？通用型人工智能存在哪些风险？针对这些风险有哪些缓解技术？

11 月 25 日，报告发布第二版关键更新，探讨了通用人工智能风险管理技术方法的发展，从训练模型拒绝有害请求到添加水印等⁶。



2 月

4. OECD 发布 AI 时代数据获取与共享政策建议

摘要：2025 年 2 月，经济合作与发展组织发布《AI 时代下增强数据获取和共享的政策建议》⁷，为各国应对 AI 时代数据短缺问题提供指导。建议强调了政府如何提升对数据及某些人工智能模型的访问

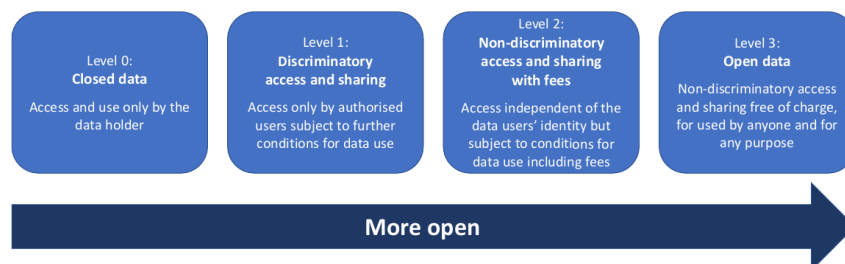
⁵ https://internationalaisafetyreport.org/sites/default/files/2025-10/international_ai_safety_report_2025_executive_summary_chinese.pdf.

⁶ <https://internationalaisafetyreport.org/sites/default/files/2025-12/second-key-update-english.pdf>.

⁷ OECD. https://www.oecd.org/en/publications/enhancing-access-to-and-sharing-of-data-in-the-age-of-artificial-intelligence_23a70dca-en.html

与共享，同时保障隐私及知识产权等其他权利和利益。建议提出数据开放图谱概念，将数据获取共享分为三个等级，强调采用隐私增强技术，促进数据在组织间的可查找性、可访问性、互操作性和可重用性。该建议为全球 AI 治理重要的数据治理基础提供了权威框架。

Figure 1. The degrees of data openness



5. 《ALECSO 人工智能伦理宪章》起草委员会在 ALECSO 总部召开会议

摘要：2025 年 2 月 4-7 日，《ALECSO 人工智能伦理宪章》起草委员会于在突尼斯 ALECSO 总部召开会议。会议期间，委员会开始着手制定宪章第一版草案，这是建立全面框架以促进阿拉伯国家负责任且合乎道德使用的重要一步。会议吸引了技术、伦理和法律领域的著名阿拉伯专家，从而确保宪章涵盖所有与人工智能相关的方面，并与地区需求最契合。宪章草案包含教育、文化和科学研究三个主要重点领域，该草案有助于统一阿拉伯在人工智能伦理领域的努力，并加强该地区和国际在这一关键领域的合作⁸。

6. OECD 推出全球框架，监测 G7 广岛 AI 行为准则实施情况

摘要：2025 年 2 月 7 日，经合组织（OECD）发布了全球首个企业报告框架，以促进企业对其推广安全、可靠人工智能所作的努力进行透明化报告。该框架同时用于监测《广岛进程国际人工智能行为准

⁸ Arab League Educational, Cultural and Scientific Organization.

<https://www.alecso.org/nsite/en/newscat/4673-%d8%a7%d8%ac%d8%aa%d9%85%d8%a7%d8%b9-%d9%84%d8%ac%d9%86%d8%a9-%d8%b5%d9%8a%d8%a7%d8%ba%d8%a9-%d9%85%d9%8a%d8%ab%d8%a7%d9%82-%d8%a7%d9%84%d8%a3%d9%84%d9%83%d8%b3%d9%88-%d9%84%d8%a3%d8%ae%d9%84%d8%a7%d9%82%d9%8a%d8%a7%d8%aa-%d8%a7%d9%84%d8%b0%d9%83%d8%a7%d8%a1-%d8%a7%d9%84%d8%a7%d8%b5%d8%b7%d9%86%d8%a7%d8%b9%d9%8a-%d8%a8%d9%85%d9%82%d8%b1-%d8%a7%d9%84%d9%85%d9%86%d8%b8%d9%85%d8%a9-2>

则》的落实情况。企业将首次能够提供在人工智能风险管理方面的可比信息，包括风险评估、事件报告及信息共享机制等行动与实践，从而推动先进人工智能系统研发过程中的信任与问责。全球多家领先的先进人工智能系统开发者参与了此项目，并在试点阶段发挥了关键作用，共同测试框架功能并确保其有效性。包括亚马逊、Anthropic、富士通、谷歌、KDDI、微软、NEC、NTT、OpenAI、Preferred Networks Inc.、乐天集团、Salesforce 和软银在内的众多人工智能领域重要企业已承诺完成该框架的首轮申报⁹。

7. 巴黎 AI 行动峰会发布全球治理宣言

摘要：2025 年 2 月 10-11 日，法国和印度共同主持的 AI 行动峰会在巴黎举行。此次峰会汇聚了来自世界各地的国家元首和政府首脑、国际组织领导人、企业界的首席执行官、学术界代表、非政府组织、艺术家以及民间社会成员等多方力量。经过两天的深入讨论与交流，共有 62 个国家及非盟和欧盟共同签署了《关于发展包容、可持续的人工智能造福人类与地球的声明》，以推动人工智能技术的健康发展。然而，美国和英国因主张“维护技术领导权”和“反对意识形态束缚”而选择不签署该宣言¹⁰。

3 月

8. 中国出台《人工智能生成合成内容标识办法》

摘要：2025 年 3 月 7 日，中国国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局四部门联合发布《人工智能生成合成内容标识办法》。2025 年 9 月 1 日起全面实施。办法要求所有 AIGC 服务在文本、图片、音频、视频等元数据中嵌入“隐式标识”，并在用户端以“显式标识”提示内容由 AI 生成。违规企业将面临暂

⁹ OECD. <https://www.oecd.org/en/about/news/press-releases/2025/02/oecd-launches-global-framework-to-monitor-application-of-g7-hiroshima-ai-code-of-conduct.html>.

¹⁰ ELYSEE. <https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet>.

停服务、罚款乃至吊销牌照的处罚¹¹。

9. 西班牙批准《人工智能法草案》并启动监管沙盒

摘要：西班牙部长理事会于 2025 年 3 月 11 日批准了《西班牙人工智能法草案》初稿。2025 年 4 月，12 个人工智能项目被选中参与研发沙盒，结果将用于发布一份公开的良好实践报告，指导未来国家人工智能监管¹²。

4 月

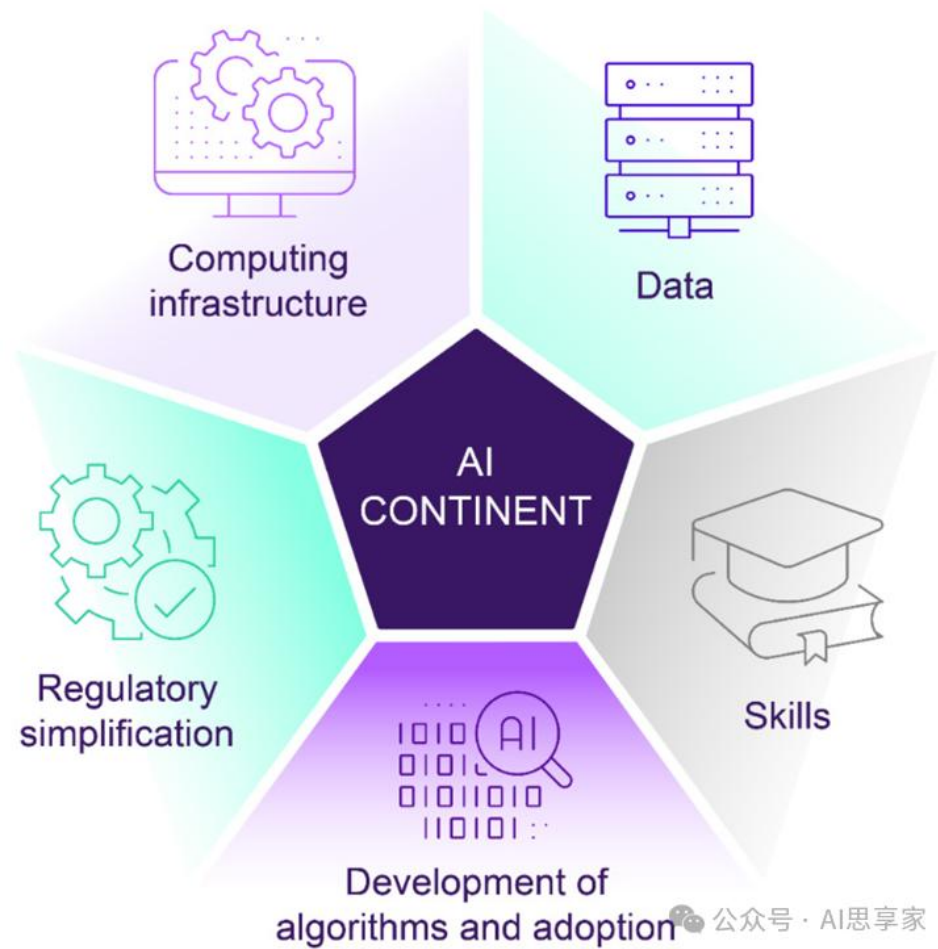
10. 欧盟发布人工智能大陆行动计划

摘要：2025 年 4 月 9 日，欧盟委员会发布《人工智能大陆行动计划》(AI Continent Action Plan)¹³。该计划提到，为了使欧盟成为一个“人工智能大陆”，必须在算力基础设施、数据、应用、人才、监管五个关键领域加快并加强努力。该计划强调欧盟机构、各国政府、企业、研究人员和开发者需紧密合作，欧盟人工智能办公室将与成员国通过人工智能委员会密切协作，确保政策一致性和适应技术发展。欧盟将通过双边和多边合作，推动安全、可信和以人为本的人工智能发展。

¹¹ 中华人民共和国国家互联网信息办公室. https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm.

¹² White & Case. <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-spain>.

¹³ European Commission. <https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan>.



11. 非洲联盟以本土化 AI 战略赋能非洲发展与转型

摘要: 2025 年 4 月，在非洲全球人工智能峰会（Global AI Summit on Africa）关于人工智能时代领导力的高级小组讨论会上，非洲联盟委员会强调了致力于将非洲定位为全球人工智能革命领导者的承诺。其核心举措是已获批准的《大陆人工智能战略：转型路线图》及五年实施计划，旨在指导成员国将人工智能纳入国家发展议程，聚焦治理、农业、医疗等关键领域，并通过留住和吸引人才、公私合作、法规建设等措施构建包容可持续的 AI 生态。同时，联盟强调 AI 需与非洲大陆自贸区协同，促进数字贸易，并重视开发体现非洲语言文化多样性的 AI 解决方案，以确保技术包容性与数字主权，最终助力非洲实现社会经济转型¹⁴。

¹⁴ African Union. <https://au.int/en/pressreleases/20250404/african-union-champions-ai-africas-socioeconomic-transformation>.

12. IEEE 智能家居多模态 AI 标准立项

摘要：2025 年 4 月，由中国海信视像牵头制定的《智能家居场景下多模态大模型技术要求》IEEE 标准正式立项¹⁵。标准涵盖**数据处理、模型训练、交互安全**等全链条要求，解决了设备协同性不足、用户体验不一致等行业痛点。作为**全球首个聚焦智能家居领域的多模态大模型技术要求**，该标准为人工智能技术与智能家居场景的深度融合建立了技术坐标系，标志着我国在智能家居 AI 核心技术领域实现从技术输出到标准引领的跨越。

13. 沙特阿拉伯发布《全球人工智能枢纽法案》

摘要：2025 年 4 月 14 日，沙特阿拉伯通信、空间与技术委员会（CSTC）发布了《全球人工智能枢纽法案》的首份草案，旨在通过设立“主权数据中心”（即“数据使馆”）吸引外国政府与私营企业在该国发展及部署先进技术。草案规定了**三类枢纽：私人枢纽**（数据完全受外国法律管辖）、**扩展枢纽**（允许私营运营商参与）和**虚拟枢纽**（由服务商提供虚拟托管）。该法律赋予沙特内阁出于国家安全等原因终止协议的权利，目前正公开征求意见。此举旨在提升沙特作为国际技术与数据合作枢纽的竞争力¹⁶。

14. 中国出台全球首部气象 AI 专项法规

摘要：2025 年 4 月 23 日，中国气象局与国家互联网信息办公室联合发布《人工智能气象应用服务办法》，要求 AI 气象服务需通过算法备案、安全评估，并对生成内容进行标识¹⁷。该管理办法为**世界气象组织成员国中首个 AI 细分领域法规**，规范了数据开放、模型研发及国际合作等环节，推动 AI 在灾害预警、数值预报中的安全应用。

¹⁵ 光明科技. https://tech.gmw.cn/2025-04/24/content_37987561.htm.

¹⁶ LEXOLOGY. <https://www.lexology.com/library/detail.aspx?g=0b19e138-7d26-4ec4-939a-207c01ceb028>.

¹⁷ 中国气象局. https://www.cma.gov.cn/zfxxgk/zc/gz/202504/t20250429_7036854.html.

15. 中央政治局就人工智能发展与监管进行集体学习 习近平总书记强调推动人工智能健康有序发展

摘要：2025年4月25日下午，中共中央政治局就加强人工智能发展和监管进行第二十次集体学习。习近平总书记在主持学习时强调，面对新一代人工智能技术快速演进的新形势，要充分发挥新型举国体制优势，坚持自立自强，突出应用导向，推动我国人工智能朝着有益、安全、公平方向健康有序发展¹⁸。

16. 欧盟发布生成式 AI 研究使用指南，为科研人员提供行动方向

摘要：2025年4月，欧盟委员会发布了《关于在研究中负责任使用生成式人工智能的动态指南》（第二版）（*Living guidelines on the responsible use of generative AI in research*）¹⁹。该指南为科研人员、研究机构以及研究资助机构提供了清晰的行动方向，为生成式 AI 在研究中的负责任使用提供了指导，有助于确保研究诚信，并推动科学研究的进步。

5月

17. 美国废止“AI 扩散规则”并升级芯片管制

摘要：2025年5月13日，美国商务部宣布废止拜登时期“人工智能扩散出口管制框架”，同时发布三份配套指引，继续严控对华高端 AI 芯片出口，这三份文件分别是《关于可能适用于先进计算芯片及其他用于训练 AI 模型商品的管制的政策声明》《关于通用禁令 10（GP10）对中华人民共和国（PRC）先进计算芯片适用的指南》《关于防止先进计算芯片转移的行业指南》²⁰。新规引入“芯片位置验证”与“制造子系统”管制，并威胁动用“外国直接产品规则”迫使盟友

¹⁸ 求是网. <https://www.qstheory.cn/20250508/c623cf72088740f890c8801d127773e8/c.html>.

¹⁹ European Commission. https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf.

²⁰ Bureau of Industry and Security. <https://www.bis.gov/press-release/department-commerce-announces-rescission-biden-era-artificial-intelligence-diffusion-rule-strengthens>.

对齐出口限制。政策凸显特朗普政府“轻国内监管、重出口封锁”的 AI 治理转向。

18. 金砖国家人工智能高级别论坛在巴西利亚举行，《金砖国家人工智能展望报告（2025）》发布

摘要：2025 年 5 月 20 日，金砖国家人工智能高级别论坛在金砖轮值主席国巴西首都巴西利亚举办。本次论坛是 2025 年金砖国家工业部长会议系列活动之一，以“智汇金砖 AI 创未来”为主题，由中国工业和信息化部，巴西公共服务管理和创新部，巴西发展、工业、贸易和服务部与巴西科技创新部共同主办，中国—金砖国家人工智能发展与合作中心承办。会上发布了《金砖国家人工智能展望报告（2025）》《人工智能典型产品与应用案例集》等成果²¹。

19. 日本首部人工智能专项法律出台

摘要：2025 年 5 月 28 日，日本参议院通过《人工智能相关技术研究开发及应用推进法》，设立由首相牵头的“AI 战略本部”，统筹制定“AI 基本计划”。法案首次明确企业需配合政府调查 AI 滥用行为（如虚假信息生成、著作权侵权），违规者可能被公开名称。该法平衡创新与风险，强调通过国际标准协同提升日本 AI 产业竞争力²²。该法案于 2025 年 9 月 1 日全面实施²³。

6 月

20. 七国集团领导人发布关于人工智能促进繁荣的声明

摘要：2025 年 6 月 17 日，七国集团领导人发布关于人工智能促进繁荣的声明。声明指出，七国集团将利用现有的政府人工智能专业知识，建立七国集团人工智能网络（GAIN），推动重大挑战；制定推广成功人工智能项目的路线图；并为会员创建开源且可共享的 AI 解

²¹ 中华人民共和国工业和信息化部。

https://wap.miit.gov.cn/xwfb/bldhd/art/2025/art_b43c3e8cc47c40a395be979fa87ee64e.html.

²² 新华网。<https://www3.xinhuanet.com/20250528/bbf21c32bc4043cbb39a02e36e6e55ab/c.html>.

²³ Prime Minister's Office of Japan. <https://www.kantei.go.jp/jp/103/actions/202509/12jinkoutchinou.html>.

决方案目录。加速公共部门人工智能的转型采用，包括大语言模型和数字基础设施。声明承诺加速公共部门 AI 应用、支持中小企业 AI 采用、应对 AI 能源挑战、与新兴市场合作等²⁴。

21. 第三届联合国教科文组织全球人工智能伦理论坛在泰国曼谷举行

摘要：2025 年 6 月 24-27 日，第三届联合国教科文组织全球人工智能伦理论坛在泰国曼谷举行。本次论坛重点展示自 2021 年联合国教科文组织《人工智能伦理建议书》通过以来取得的成就，并强调通过切实可行的行动持续推进人工智能伦理建设的必要性²⁵。

7 月

22. 金砖国家元首签署“人工智能全球治理宣言”

摘要：2025 年 7 月 6 日，在第十七次金砖国家领导人会晤期间，通过了《金砖国家领导人关于人工智能全球治理的声明》（*BRICS Leaders' Declaration on Global Governance of Artificial Intelligence*）²⁶。该声明核心为金砖国家主张建立以联合国为核心的包容性人工智能全球治理框架，强调数字主权与发展权，反对治理碎片化。呼吁确保发展中国家公平获取 AI 技术、数据与基础设施，提倡开放创新与跨利益相关方协作，同时关注伦理、安全、偏见等风险，推动 AI 服务于可持续发展、环境保护、劳动保障及教育转型，旨在为全球南方国家争取更多话语权与利益。

23. 中国在 2025 世界人工智能大会推出人工智能治理与安全系列倡议

摘要：2025 年 7 月 26 日，2025 世界人工智能大会暨人工智能全球治理高级别会议发表《人工智能全球治理行动计划》²⁷，提出 13 条

²⁴ G7 Research Group. <https://g7.utoronto.ca/summit/2025kananaskis/250617-ai.html>.

²⁵ UNESCO. <https://www.unesco.org/en/forum-ethics-ai>.

²⁶ https://brics.br/en/documents/presidency-documents/250706_brics_ggai_declarationfinal.pdf/@@download/file.

²⁷ 中华人民共和国中央人民政府. https://www.gov.cn/yaowen/liebiao/202507/content_7033929.htm.

具体举措，涵盖共同把握 AI 机遇、促进创新发展、赋能千行百业、数字基础设施建设等。该行动计划强调向善为民、尊重主权、发展导向、安全可控等原则。

与此同时，在大会“人工智能发展与安全”全体会议上，中国信息通信研究院、中国人工智能产业发展联盟、清华大学、上海人工智能实验室、中国电子信息产业发展研究院等机构联合发布了《中国人工智能安全承诺框架》。该框架在 2024 年 12 月版《人工智能安全承诺》基础上，进一步纳入了加强人工智能安全治理国际合作、防范前沿人工智能安全风险等新内容，彰显了中国产业界愿与全球协同共治、推动人工智能向善发展的开放态度与坚定承诺²⁸。

24. 欧盟发布《通用人工智能行为准则》

摘要：2025 年 7 月 10 日，欧盟委员会发布《通用人工智能行为准则》（*The General-Purpose AI Code of Practice*）²⁹，作为对《人工智能法案》的重要补充。准则分为透明度、版权保护、安全与保障三个章节，要求通用 AI 模型开发者制定并持续更新模型文档，建立版权政策，识别系统性风险。该准则为 2025 年 8 月 2 日生效的通用 AI 规则提供了具体操作指导。

25. 欧盟委员会正式发布《通用人工智能模型提供商指南》

摘要：2025 年 7 月 18 日，欧盟委员会正式发布《通用人工智能模型提供商指南》³⁰（以下简称《指南》），对《人工智能法案》相关义务的范围作出了权威解释。《指南》于 2025 年 8 月 2 日正式生效，直接适用于所有计划在欧盟市场投放的通用人工智能模型（GPAI Model）。其核心价值在于为产业链各方提供稳定可预期的合规框架，使开发者能够在清晰的规则基础上持续开展创新。《指南》还明确了

²⁸ <https://www.cww.net.cn/article?id=602676>.

²⁹ European Commission. <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>.

³⁰ European Commission. <https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act>.

“通用人工智能模型”的定义，即训练算力超过 10^{23} FLOP、能够生成文本、语音、图像或视频的模型，该类模型通常被认定为通用 AI 模型。

说明：《通用人工智能模型提供商指南》与《通用人工智能行为准则》共同构成了欧盟对通用人工智能（GPAI）监管的实施框架，但侧重点和性质不同。前者作为官方解释性文件界定了“是什么”和“谁负责”，后者作为自愿性操作手册指导“怎么做”。

26. 国际电信联盟“AI 向善全球峰会”推动国际治理新倡议

摘要：2025 年 7 月 7-11 日，国际电信联盟在瑞士日内瓦举办信息社会世界峰会二十年审议（WSIS+20）高级别活动、“人工智能向善”全球峰会等系列活动。150 家企业、1.5 万名代表参加了“人工智能向善全球峰会”。峰会发布“人工智能技能联盟”“全球粮食系统人工智能倡议”“人工智能标准交换数据库”等多项治理工具，旨在帮助发展中国家制定本地标准、培养人才并应对深度伪造。会议成果为后续国际条约谈判奠定基础。我国工业和信息化部副部长单忠德出席相关活动并作主旨发言，介绍了中国人工智能开源创新情况，提出要积极构建高效协同、包容互信、开放合作、普惠发展的创新生态³¹。

8 月

27. 欧盟《人工智能法案》部分条款正式生效

摘要：2025 年 8 月 2 日，欧盟《人工智能法案》关于通用人工智能模型的关键条款正式生效，要求开发者披露训练数据来源、遵守版权保护规则，并建立系统性风险评估机制。这是全球首部综合性 AI 监管法规进入强制执行阶段的重要里程碑，标志着 AI 治理从原则走向具有法律约束力的具体规则，对全球 AI 产业发展产生深远影响³²。

³¹ 中华人民共和国工业和信息化部。

https://www.miit.gov.cn/xwfb/bldhd/art/2025/art_67d2ae918fec4762ada88910964698bc.html.

³² European Commission. <https://digital-strategy.ec.europa.eu/en/news/eu-rules-general-purpose-ai-models-start-apply-bringing-more-transparency-safety-and-accountability>.

28. 中国国务院印发《关于深入实施“人工智能+”行动的意见》

摘要：2025年8月，国务院印发《关于深入实施“人工智能+”行动的意见》³³。《意见》明确了“人工智能+”行动的六大重点领域，包括科学技术、产业发展、消费提质、民生福祉、治理能力以及全球合作。为夯实发展基础，《意见》围绕提升模型基础能力、加强数据供给创新、强化智能算力统筹、优化应用发展环境、促进开源生态繁荣、加强人才队伍建设、完善政策法规保障、提升安全能力水平等方面，系统部署了基础支撑能力的建设路径。

《意见》提出阶段性发展目标：到2027年，人工智能将在六大重点领域实现广泛深度融合，新一代智能终端、智能体等应用普及率预计超过70%；到2030年，人工智能将全面赋能我国高质量发展，相关应用普及率超过90%，智能经济成为经济发展的重要增长极，推动技术普惠与成果共享；到2035年，我国将全面迈入智能经济与智能社会新阶段，为基本实现社会主义现代化提供坚实支撑。

29. 联合国拉丁美洲及加勒比经济委员会发布《加强加勒比地区人工智能准备度》报告

摘要：8月25日，联合国拉丁美洲及加勒比经济委员会发布《加强加勒比地区人工智能准备度》报告³⁴，分析了人工智能在加勒比地区应用的潜力与风险，并提出提升区域人工智能治理能力的政策建议，以支持可持续发展目标的实现。该报告指出，加勒比地区在利用人工智能促进可持续发展方面潜力巨大，但面临显著挑战。地区各国虽积极参与全球AI治理对话，但国家层面的战略、法律与基础设施仍处于起步阶段，存在数字鸿沟、数据成本高昂、5G网络部署困难等问题。同时，数字与AI技能基础薄弱，且女性在高等教育IT领域代表性不足。AI发展还需应对环境可持续性挑战，如电子废物管理和能

³³ 中华人民共和国国务院新闻办公室. http://www.scio.gov.cn/zdgz/jj/202509/t20250901_928364.html.

³⁴ <https://repositorio.cepal.org/server/api/core/bitstreams/6658d38e-c635-4e82-9c0c-d5802d446e405/content>.

源需求。报告建议通过加强国际合作、制定综合政策、投资基础设施与技能培训、促进性别平等及推行循环经济，以系统提升该地区的 AI 准备度³⁵。

30. 联合国设立“AI 独立国际科学小组”与全球对话机制

摘要：2025 年 8 月 26 日，联合国大会通过决议，设立由 40 名独立专家组成的“人工智能独立国际科学小组”及“全球人工智能治理对话”常设论坛，2026 年起每年发布风险报告并召开政府间会议，旨在加强全球人工智能治理，支持可持续发展目标，缩小数字鸿沟³⁶。

9 月

31. ISO 发布全球首个 AI 可解释性国际标准

摘要：2025 年 9 月，国际标准化组织（ISO）与 IEC 联合发布 TS 6254:2025 标准《信息技术 人工智能 机器学习模型与人工智能系统可解释性与可理解性的目标及方法》³⁷。该标准作为全球范围内首个系统性地为 AI 可解释性提供框架与指南的国际文件，首次明确“可解释性”与“可理解性”的定义及技术路径。标准覆盖 AI 全生命周期，针对开发者、用户、监管机构等 11 类利益相关方提出差异化要求，例如审计方需评估模型决策逻辑，受影响个体有权获取特定决策解释，为全球 AI 透明化提供统一框架³⁸。

32. EDPB 发布首份《数字服务法》（DSA）与《通用数据保护条例》（GDPR）协同适用指南

摘要：2025 年 9 月 12 日，欧洲数据保护委员会（EDPB）通过首份关于《数字服务法》（DSA）与《通用数据保护条例》（GDPR）协同适用的指南³⁹。该指南旨在协调两部法律在实践中的适用，明确

³⁵ ECLAC. <https://www.cepal.org/en/publications/82423-strengthening-artificial-intelligence-readiness-caribbean-policy-brief>.

³⁶ United Nations. <https://press.un.org/en/2025/ga12699.doc.htm>.

³⁷ ISO. <https://www.iso.org/standard/82148.html>.

³⁸ 中华人民共和国商务部. <https://chinawto.mofcom.gov.cn/article/jsbl/dtxx/202510/20251003604456.shtml>.

³⁹ European Data Protection Board (EDPB). https://www.edpb.europa.eu/news/news/2025/interplay-between-dsa-and-gdpr-edpb-adopts-guidelines_en.

当 DSA 涉及个人数据处理时，应如何依据 GDPR 的相关定义与原则执行，以提升法律适用的连贯性与确定性。指南重点涵盖通知与行动机制、平台推荐系统、未成年人保护、广告透明度及基于特殊数据类别画像的广告禁止等多个交叉领域。此外，EDPB 还就不同监管机构间的协作提出实务指引。未来将推动 GDPR 与《数字市场法》《人工智能法》等新数字法律的衔接。

33. 北京发布《从人工智能到超级智能北京宣言》

摘要：2025 年 9 月 14 日，在中国国际服务贸易交易会期间，第三届人工智能国际治理学术论坛（IAIGF）暨第五届智能+产业融合高端智库论坛（IHETT）在北京举办。论坛以“从人工智能（AI）到超级智能（ASI）”为主题，汇聚了全球院士专家、国际组织代表及产业界人士。论坛的重要成果是正式发布了《从人工智能到超级智能北京宣言》。该宣言前瞻性地针对未来超级智能的发展，提出了“可信、可控、共治、普惠、绿色、共生”六大核心原则，并规划了涵盖基础研究与技术攻关、基础设施建设、重点场景应用、产业生态与人才建设四大领域的二十项具体行动⁴⁰。

34. 中国发布《人工智能安全治理框架》2.0 版

摘要：9 月 15 日，在 2025 年国家网络安全宣传周主论坛上，《人工智能安全治理框架》2.0 版正式发布。作为全国网安标委技术文件，《框架》2.0 版在 2024 年《框架》基础上，结合人工智能技术发展和应用实践，持续跟踪风险变化，完善优化风险分类，研究探索风险分级，动态调整更新防范治理措施。有助于推进多边机制下人工智能安全治理合作，推动世界范围内技术成果的普惠共享，确保人类社会共享人工智能发展的红利⁴¹。

⁴⁰ 央视网. <https://gongyi.cctv.com/2025/09/19/ARTILMuYEldXnxmaGf2yevIB250919.shtml>.

⁴¹ 中华人民共和国国家互联网信息办公室. https://www.cac.gov.cn/2025-09/15/c_1759653448369123.htm.

安全风险与技术应对措施、综合治理措施映射表

安全风险		技术应对措施	综合治理措施	
人工智能技术内生安全风险	模型算法安全风险	可解释性不足	4.1.1 (a)	<ul style="list-style-type: none"> 提升研发应用全生命周期安全能力 建设人工智能安全测评体系 完善数据安全和个人信息保护规范
		偏见、歧视	4.1.1 (b)	
		鲁棒性不强	4.1.1 (c)	
		输出决策不可靠	4.1.1 (a) (b) (c)	
		外部对抗攻击	4.1.1 (c)	
	模型缺陷扩散	4.1.1 (d)		
	数据安全风险	违规收集使用数据	4.1.2 (a)	
训练数据内容不当		4.1.2 (b) (c) (d) (e)		
训练数据标注不规范		4.1.2 (c)		
数据和个人信息泄露		4.1.2 (d)		
人工智能技术应用安全风险	网络系统安全风险	组件和算力安全	4.2.1 (b) (c) (d) (e)	<ul style="list-style-type: none"> 强化开源生态安全和供应链安全 实施应用分类及安全风险分级管理 推广人工智能生成合成内容可追溯管理 安全有效释放重要行业应用需求 共享人工智能安全风险威胁信息
		网络暴露面扩大	4.2.1 (a) (b) (c) (d)	
		供应链安全	4.2.1 (c) (d)	
		网络攻击滥用	4.2.1 (a) (e)	
	信息内容安全风险	输出违法有害信息	4.2.2 (a) (b) (c)	
		混淆事实、误导用户	4.2.2 (c)	
		污染网络内容生态	4.2.2 (a) (b) (c)	
	现实安全风险	经济社会运行安全的新挑战	4.2.3 (a) (b) (c) (d) (e)	
		被违法犯罪活动利用“作恶”	4.2.3 (a) (b) (c) (d) (e)	
		核生化导武器知识、能力失控	4.2.3 (a) (b) (c) (d) (e)	
	认知安全风险	加剧“信息茧房”效应	4.2.4 (a) (b) (c)	
助力开展认知战		4.2.4 (a) (b) (c)		
人工智能应用衍生安全风险	社会和环境安全风险	冲击劳动就业结构	4.3.1 (a)	<ul style="list-style-type: none"> 建立健全人工智能安全法律法规 构建人工智能科技伦理准则 增进协同应对人工智能失控风险的共识 加大人工智能安全人才培养力度 提升全社会的人工智能安全意识 促进人工智能安全治理国际交流合作
		挑战资源供需平衡	4.3.1 (a) (b)	
	伦理安全风险	加剧社会偏见、扩大智能鸿沟	4.3.2 (a) (b) (c)	
		冲击教育、抑制创新	4.3.2 (a) (b) (c)	
		加剧科研伦理风险	4.3.2 (a) (b) (c)	
		拟人化交互的沉迷依赖	4.3.2 (a) (c)	
		挑战现行社会秩序	4.3.2 (a) (b) (c)	
		“自我意识”觉醒、脱离人类控制	4.3.2 (a) (b) (c)	

35. 2025 中国—东盟人工智能部长圆桌会议在广西南宁召开

摘要：9月18日，国家发展改革委、广西壮族自治区人民政府联合主办的2025中国—东盟人工智能部长圆桌会议在广西南宁召开。圆桌会宣布在广西启动建设中国—东盟国家人工智能应用合作中心，围绕夯实人工智能发展基础、提供开源开放服务、加强产业合作对接、促进人才培育等方面开展具体合作，推动务实合作项目落地，共促人工智能技术普惠应用。此外，圆桌会提出了开展中国—东盟“人工智能+”行动，提出共建中国—东盟人工智能部长合作机制，并发布了

一批重要合作成果⁴²。

36. 美英签署《科技繁荣协议》

摘要：2025年9月18日，美国总统特朗普与英国首相斯塔默签署了《科技繁荣协议》⁴³，旨在建立全面的人工智能合作伙伴关系。核心内容包括：建立 AI 安全与标准协同机制，推动双方顶尖安全研究机构在模型评估与标准制定上深入合作；启动联合旗舰科研计划，聚焦“AI for Science”，在生物技术、精准医学等领域共同攻坚；深化太空 AI 合作，支持月球与火星探索等任务。此外，协议致力于共同构建有利于创新的 AI 政策框架，促进以美英技术为主导的 AI 全栈解决方案出口，并加强在 AI 基础设施、硬件创新及未来人才培养方面的协作⁴⁴。

37. 美国加州通过《前沿人工智能透明法案》

摘要：2025年9月29日，加州议会通过《前沿人工智能透明法案》(TFAIA)，首次将“训练算力 $\geq 10^{26}$ FLOPs”的大模型定义为“前沿模型”，强制开发者披露透明度报告、安全事件及风险评估。法案把微调、强化学习等后续训练全部纳入算力门槛，防止企业通过“分阶段训练”规避监管。该立法被视为美国州层面对超大模型实施“算力红线”监管的里程碑⁴⁵。

10 月

38. 欧盟委员会发布《应用人工智能战略》

摘要：2025年10月8日，欧盟委员会正式发布《应用人工智能战略》(Apply AI Strategy)。欧盟《应用人工智能战略》旨在将欧盟打

⁴² 中华人民共和国国家发展和改革委员会。

https://www.ndrc.gov.cn/fzggw/wld/wcl/zyhd/202509/t20250918_1400573.hptml.

⁴³ The White House. <https://www.whitehouse.gov/articles/2025/09/president-trump-signs-technology-prosperity-deal-with-united-kingdom/>.

⁴⁴ The White House. <https://www.whitehouse.gov/presidential-actions/2025/09/memorandum-of-understanding-between-the-government-of-the-united-states-of-america-and-the-government-of-the-united-kingdom-of-great-britain-and-northern-ireland-regarding-the-technology-prosperity-de/>.

⁴⁵ California Legislative Information.

https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB53&search_keywords=Transparency+in+Frontier+AI+Act.

造为“人工智能大陆”，提升战略行业竞争力和技术主权，重点推动中小企业应用与创新。该战略倡导“AI先行”政策，鼓励在决策中优先评估AI方案，并推广公共部门优先采用欧洲本土及开源AI解决方案。其主要内容围绕三大支柱展开：在11个关键行业与公共部门实施旗舰计划；通过建设人工智能工厂、测试设施与创新中心等支持措施应对共同挑战；建立由应用人工智能联盟和AI观察站组成的新治理体系，确保政策贴合实际需求。该战略与《科学人工智能战略》及《数据联盟战略》协同推进，共同构建欧洲AI生态⁴⁶。

39. 欧盟委员会发布《欧洲科学人工智能战略》

摘要：2025年10月8日，欧盟委员会发布《欧洲科学人工智能战略》（*European Strategy for Artificial Intelligence in Science*），该战略的虚拟中枢与核心引擎，是“欧洲科学AI资源平台（RAISE）”计划，它旨在打破国家与机构的壁垒，将全欧的顶尖算力、高质量数据、科研经费和卓越人才汇入一个协同网络。RAISE具有双重使命，一是“Science for AI”，攻关AI前沿技术本身，开发下一代AI。二是“AI in Science”，推动AI在各科学领域的广泛应用，催生颠覆性发现⁴⁷。

40. 中国国家互联网信息办公室发布《政务领域人工智能大模型部署应用指引》

摘要：2025年10月10日，国家互联网信息办公室发布《政务领域人工智能大模型部署应用指引》，为各级政务部门开展大模型部署应用提供了基本遵循与实施导向。文件聚焦**政务服务、社会治理、机关办公和辅助决策**四大类型，明确了13种典型应用场景，并对政务大模型的规范部署、运行管理与保障措施提出了具体要求⁴⁸。

⁴⁶ European Commission. <https://digital-strategy.ec.europa.eu/en/policies/apply-ai>.

⁴⁷ European Commission. https://research-and-innovation.ec.europa.eu/document/clafd7d0-ff65-4f84-be48-b0e0949596c5_en.

⁴⁸ 中华人民共和国国家互联网信息办公室. https://www.cac.gov.cn/2025-10/10/c_1761819469929310.htm.

41. 美日签署《科技繁荣协议》

摘要：2025年10月28日，美日签署《科技繁荣协议》⁴⁹，旨在推动人工智能、量子技术和生物技术领域的战略合作，人工智能被明确列为双方合作的战略重心，具体涵盖八大方向。这一协议建立在今年7月日本承诺向美国投入5500亿美元的基础上⁵⁰。在10月的具体落实中，AI基础设施成为投资重点，日本及多家日本公司明确承诺开展对美投资。

11月

42. 欧盟启动 AI 生成内容标记准则制定

摘要：2025年11月5日，欧盟委员会正式启动了关于AI生成内容标记与标签行为准则的制定工作。这一举措基于《人工智能法案》的透明度要求，旨在应对日益增长的虚假信息和欺诈风险。该准则将作为自愿性工具，帮助生成式AI系统的提供者和部署者有效履行其透明度义务，支持包括合成音频、图像、视频和文本在内的AI生成内容以机器可读格式进行标记⁵¹。欧盟启动《AI生成内容透明度实践准则》制定，是《AI法案》从法律文本走向现实落地的关键一步。

43. 欧盟《数字综合法案》提案引发争议

摘要：2025年11月19日，欧盟委员会发布《数字综合法案》提案，拟修订AI规则、网络安全、隐私框架等五方面内容，包括简化GDPR合规要求、更新Cookie规则⁵²。尽管旨在促进数字经济发展，但被批评“系统性削弱隐私保护”，民间组织NOYB指出其可能侵犯个人数据权利，且缺乏欧盟内部共识。该提案将进入欧洲议会审议阶段，可能影响全球AI企业在欧盟的运营合规成本^{53,54}。

⁴⁹ The White House. <https://www.whitehouse.gov/articles/2025/10/u-s-japan-technology-prosperity-deal/>.

⁵⁰ The White House. <https://www.whitehouse.gov/fact-sheets/2025/10/28195/>.

⁵¹ European Commission. <https://digital-strategy.ec.europa.eu/en/news/commission-launches-work-code-practice-marking-and-labelling-ai-generated-content>.

⁵² European Commission. <https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal>.

⁵³ 欧洲时报. <https://www.oushinet.com/static/content/europe/other/2025-11-20/1441121403327780910.html>.

⁵⁴ 安全内参. <https://www.secrss.com/articles/85463>.

44. 美国特朗普政府启动“创世纪计划”

摘要: 11月24日,美国总统特朗普签署行政令,启动国家级AI科研旗舰项目“创世纪计划”(Genesis)⁵⁵。这是一次对标“阿波罗计划”、“曼哈顿计划”的战略动员,代表了历史性的资源动员,旨在将AI全面融入科学研究。能源部将整合国家实验室超算与联邦科研数据,构建闭环AI实验平台,重点服务生物、量子、聚变等战略领域。计划要求十年内提升联邦科研经费产出效率,并通过政企数据共享强化国家安全与经济竞争力。

45. 韩国政府发布《AI领域规制合理化路线图》

摘要: 2025年11月27日,韩国政府发布《AI领域规制合理化路线图》(首轮新产业规制合理化方案),旨在通过消除阻碍AI产业发展的不合理或过度规制,推动AI技术创新与产业发展。路线图围绕四大领域展开: **技术开发**(如数据版权与公共数据开放)、**服务应用**(如自动驾驶与AI机器人)、**基础设施**(如数据中心建设)以及**信任与安全规范**(如高影响力AI定义)。共包含67项具体任务,涉及25个部门协作,强调现场调研与企业参与,以构建灵活的规制环境,提升韩国AI产业竞争力⁵⁶。

12月

46. 越南通过了《人工智能法》

摘要: 2025年12月10日,越南国会高票通过了《人工智能法》,将于2026年3月1日生效。这是越南首部综合性人工智能法律,旨在构建“以管理促发展”的框架,平衡风险管控与创新激励。该法确立了以人为中心的原则,并依据风险高低对AI系统实行分级管理。亮点举措包括设立国家AI发展基金、推出AI代金券支持企业、建立监管沙盒,并着重推动计算基础设施、数据开放与人才培养,以提升

⁵⁵ The White House. <https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/>.

⁵⁶ <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=11632#LINK>.

国家在数字时代的竞争力⁵⁷。

47. 美国特朗普政府签署 AI 监管行政令

摘要: 2025 年 12 月 11 日，美国总统特朗普签署一项行政令，在联邦层面统一 AI 监管规则，限制各州各自制定规则。行政令要求建立“负担最小”的全国性标准，成立 AI 诉讼特别工作组，对与行政令相悖的各州 AI 法律提出异议⁵⁸。这一政策被批评者认为可能导致 AI 监管弱化，代表了美国 AI 治理理念的重要转向。

⁵⁷ Báo Điện tử Chính phủ. <https://baochinhphu.vn/viet-nam-chinh-thuc-co-luat-tri-tue-nhan-cao-ai-102251210164948585.htm>.

⁵⁸ The White House. <https://www.whitehouse.gov/fact-sheets/2025/12/fact-sheet-president-donald-j-trump-ensures-a-national-policy-framework-for-artificial-intelligence/>.