世界互联网大会
World Internet Conference

# 全球人工智能标准发展报告


世界互联网大会人工智能专业委员会
标准推进计划

2025年11月

# 标准推进计划成员及其他贡献者

## 牵头人

魏　凯　　中国信息通信研究院

约翰·希金斯　　国际人工智能治理协会

## 专家及贡献者

（按单位名称首字母排序）

**阿里巴巴（中国）有限公司**

彭骏涛、雷永强

**北京百度网讯科技有限公司**

蒋晓琳

**北京印象笔记科技有限公司**

常　诚、岳　峰

**国际电信联盟电信标准化部门**

毗湿奴·拉姆、杨晓雅

**国际人工智能治理协会**

罗布·沃萨姆

**金砖国家未来网络研究院**

林　桢

**科大讯飞股份有限公司**

乔玉平、杨彤晖、张云畅、崔　稳

**南德商品检测上海有限公司**

方　华

**奇安信科技集团股份有限公司**

王占一、安锦程、孔　坚、姜伟生

**上海诺基亚贝尔股份有限公司**

贺　敬、王　彤、顾方方

**安谋科技（中国）有限公司**

王骏超

**北京金山云网络技术有限公司**

西京京

**重庆长安汽车股份有限公司**

罗咏刚、李雪梅、付春晖、刘　玲

**华为技术有限公司**

柳嘉琪

**国网北京市电力公司**

郝佳恺、李宇婷

**京东科技信息技术有限公司**

赖南华、梁拔剑、孙　路

**联想（北京）有限公司**

李汝鑫、胡永启

**欧盟驻华标准化专家项目**

徐　斌

**荣耀终端股份有限公司**

曾勇波、赵晓娜、李辰淑、杜睿琪

**上海商汤智能科技有限公司**

吴　庚

**上海燧原科技股份有限公司**

蒋 燕、梅敬青、王思善

**维沃移动通信有限公司**

高立发、曹宇琼

**央视国际网络有限公司**

李义彪

**云安全联盟**

黄连金、叶小倩

**之江实验室**

黄 恺、张建锋、黄丹丹、蒙贵云

**中国电信股份有限公司研究院**

张 园、刘 晴、刘小欧、杨明川

**中国科学院计算技术研究所**

王煜炜

**中国人民大学**

文继荣

**中国网络空间安全协会**

王健兵、夏文辉

**深信服科技股份有限公司**

叶润国

**世界互联网大会**

梁 昊、张雪丽

**英国标准协会**

克里斯·布朗

**浙江大华技术股份有限公司**

曲翔宇

**中国移动通信集团有限公司**

江为强

**360数字安全集团**

马 琳

**中国联合网络通信集团有限公司**

刘 琪、加雄伟、杨 斌、竹梦圆

**中国社会科学院法学研究所**

周 辉

**中国信息通信研究院**

王蕴韬、谢家乐

## 编写团队

**中国信息通信研究院人工智能研究所**

呼娜英、杨 凡、郭苏敏、徐 鹏、石 霖

**世界互联网大会**

康彦荣、陈玉杰

## 联系邮箱

research@wicinternet.org


# 前言

**人工智能正成为引领第四次科技革命的核心驱动力，对全球经济、社会和治理体系产生深刻变革。**2024年，全球人工智能市场规模突破6382亿美元，预计到2030年将达到3.6万亿美元[1]，广泛赋能各领域，加速推进数字化、网络化、智能化转型。面对这一趋势，国际标准化组织（ISO）、国际电工委员会（IEC）、国际电信联盟（ITU）、电气电子工程师学会（IEEE）等国际和国家标准化组织密集启动人工智能标准研制工作，推动建立跨区域、宽领域、多主体的技术规范与治理体系。与此同时，全球主要经济体也普遍将人工智能标准化作为战略重点，加速促进标准制定，支撑产业发展。

**全球人工智能标准化的重要性日益凸显，但部分亟需关注的问题也逐渐显现。一是**技术迭代更新提速，新兴应用场景层出不穷，而现有标准体系难以及时覆盖，基础标准滞后易引发更多信息孤岛，加大成本投入。**二是**人工智能产业链条横跨算法、数据、算力、应用等多个环节，涉及政府、科研机构、企业等多元主体，利益诉求差异导致标准协调过程复杂。**三是**不同国家和地区在标准制定中的参与能力与话语权存在明显差距，影响标准的普适性与包容性，也间接扩大全球数字鸿沟。

在此背景下，本报告立足全球人工智能标准化全景，系统分析国际标准组织和主要经济体标准化行动，研判标准加快提速、互操作性与负责任的发展态势，剖析技术迭代提速、产业链条复杂、治理理念差异、南方国家缺位等带来的挑战，提出负责任的人工智能标准发展建议。报告强调，全球合作是构建包容、互联、可持续人工智能标准体系的关键路径。实现这一目标，需要国际组织、政府、产业界和科研机构携手合作，共同应对挑战，持续推动负责任人工智能标准的制定与落实。

[1] Precedence Research. 2025年至2034年人工智能市场规模、份额和趋势[EB/OL]. 2025-08-21. https://www.precedenceresearch.com/artificial-intelligence-market

# 目录

# 01 全球人工智能标准现状

## （一）国际主要标准组织标准化行动

**人工智能标准化正成为全球治理与技术协同的重要抓手，国际组织在推动技术规则统一与治理框架互认上发挥核心作用。**当前，国际标准组织呈现多元协同格局：ITU依托信息通信领域优势，构建覆盖技术、应用与治理的多层次体系，强化安全可信与全球协同；ISO和IEC以其联合技术委员会人工智能分委会（JTC 1 SC 42）为核心，打造贯穿基础、技术与管理的敏捷式标准体系，兼顾制定思路与形式的创新；IEEE以伦理治理为导向，推动P7000系列标准用于规范人工智能系统道德规范方面的问题，与培训认证衔接，构建跨领域、全栈式标准生态。

**国际电信联盟（ITU）依托信息通信领域优势，构建多层次标准推进体系，朝着安全可信与全球协作方向演进。一是建立多层次标准化架构。**ITU人工智能标准化工作主要集中在电信标准化部门（ITU-T）。ITU-T通过10个研究组（SG）开展标准制定，由焦点组（FG）推进特定应用场景的标准预研，并辅以联合协调行动（JCA）与联络小组（CG）进行统筹（见表1）。此外，在审批过程中采取两种模式：在线快速审批流程适用于前沿技术性标准；传统审批流程则适用于涉及监管的政策性标准，由成员国审查批准，以确保规范一致性。**二是覆盖技术、应用与治理全链条。**截至2025年4月，ITU-T已发布120多项相关标准[2]，主要分为三类：其一是原生人工智能技术标准，涵盖机器学习、联邦学习、云平台与生成式人工智能，主要研究工作由SG21牵头开展；其二是人工智能赋能信息通信产业的应用标准，涉及下一代通信网络、自动化网络、智慧城市、多媒体应用，以及未来网络赋能人工智能等，由SG2、SG13、SG20、SG21开展；其三是人工智能赋能社会的治理标准，主要面向公共安全、深度伪造防范与个人数据保护，由SG17牵头开展。**三是推动安全可信与全球协同。**ITU将人工智能安全可信作为下一步标准化重点，强调在伦理原则、

[2] ITU. 标准有助于为所有人释放值得信赖的人工智能机会[EB/OL]. 2025-04-01. https://www.itu.int/hub/2025/04/standards-help-unlock-trustworthy-ai-opportunities-for-all/

透明度与问责机制上形成统一规范，并通过建立模型评估、偏差缓解和安全防护等技术基准，确保相关标准具有可操作性与产业适配性。在此基础上，ITU积极倡导全球协作，与各大国际组织建立联络机制，共同推动跨区域标准对齐，避免标准碎片化；同时，鼓励各国将人工智能标准纳入政策与监管体系，实现不同治理框架互认，从而提升全球范围内负责任人工智能治理效能。

**表1 ITU-T人工智能标准化相关内部组织**

| 研究组（SG） | 焦点组（FG） | 其他 |
| --- | --- | --- |
| • SG2：业务方面<br>• SG3：经济和政策问题<br>• SG5：环境、EMF、气候行动和循环经济<br>• SG11：协议、测试和打击假冒<br>• SG12：性能、QoS和QoE<br>• SG13：未来网络<br>• SG15：交通、通道和家庭<br>• SG17：安全<br>• SG20：物联网、数字孪生和智慧城市<br>• SG21：多媒体、内容交付和有线电视 | • FG-AINN：电信网络原生人工智能<br>• FG-AI4A：数字农业人工智能与物联网<br>• FG-AI4NDM：自然灾害管理人工智能<br>• FG-AI4H：健康人工智能<br>• FG-AN：自主网络<br>• FG-AI4EE：人工智能及其他新兴技术环境效率<br>• FG-AI4AD：自动驾驶和辅助驾驶人工智能 | • JCA-AI/ML：人工智能，包括机器学习<br>• CG-AISEC-STRAT：人工智能安全信息通信策略 |

**国际标准化组织（ISO）与国际电工委员会（IEC）以JTC 1 SC 42为核心，构建从原理到方法的标准化体系。一是形成以SC 42为核心的标准化格局。**ISO和IEC在联合技术委员会JTC 1下设人工智能分委会（SC 42）统筹人工智能标准研制，辅以信息安全、网络安全与隐私保护分委会（SC 27）与可信赖工作组（WG 13）建立协作机制（见表2）。其中，SC 42专注于人工智能基础术语、系统方法论与治理，SC 27聚焦人工智能威胁与隐私保护，WG 13统一可信赖概念与术语，奠定跨委员会标准研制的话语体系。**二是构建基础、技术与管理协同的标准体系。**SC 42目前已发布37项相关标准[3]，形成了从概念、技术到应用的递进式标准化路径。其中，基础层旨在为标准化提供方法论支撑，如ISO/IEC 22989:2022《人工智能概念与术语》、ISO/IEC 23053:2022《机器学习人工智能系统框架》等；技术层聚焦安全与可信，以应对人工智能原生技术挑战，如ISO/IEC TR 24028:2020《人工智能可信度概述》、ISO/IEC TR 24029-1:2021《神经网络鲁棒性评估第1部分：概述》等；管理层面向实践落地，旨在建立内外相结合的安全体系，如ISO/IEC 42001:2023《人工智能管理体系》、ISO/IEC 42005:2025《人工智能系统影响评估》等。**三是推动制定思路与形式的双重创新。**ISO/IEC正在基于生成式人工智能与智能体等前沿场景，探索以数据、模型与应用为核心的新兴标准方向，并将可信赖特性作为横贯全生命周期的关键指标。同时，其全面采用在线标准制定模式，以数字化反馈机制缩短制定周期，并引入机器可读的智能标准，通过模块化与可组合机制加速技术标准落地，适应人工智能快速迭代与跨领域融合的趋势。

[3] ISO. ISO/IEC JTC 1/SC 42标准：人工智能[EB/OL]. 2025-08-01. https://www.iso.org/committee/6794475/x/catalogue/p/1/u/0/w/0/d/0

**表2 ISO/IEC JTC 1人工智能标准化相关内部组织**

| 人工智能分委会（SC 42） | 信息安全、网络空间安全和隐私保护分委会（SC 27） |
| --- | --- |
| • WG 1：基础标准工作组 | • WG 1：信息安全管理体系工作组 |
| • WG 2：数据工作组 | • WG 2：密码技术与安全机制工作组 |
| • WG 3：可信赖工作组 | • WG 3：安全评价、测试和规范工作组 |
| • WG 4：用例和应用工作组 | • WG 4：安全控制和服务工作组 |
| • WG 5：人工智能系统的计算方法和计算特性工作组 | • WG 5：身份管理和隐私技术工作组 |
| • JWG 2：人工智能系统测试联合工作组 | • JWG 7：生物特征识别的网络安全测试与评价 |
| • JWG 3：人工智能医疗信息联合工作组 | |
| • JWG 4：人工智能系统功能安全联合工作组 | |
| • JWG 5：自然语言处理联合工作组 | |
| • JWG 6：人工智能合格评定联合工作组 | |
| • JWG 7：人工智能金融联合工作组 | |

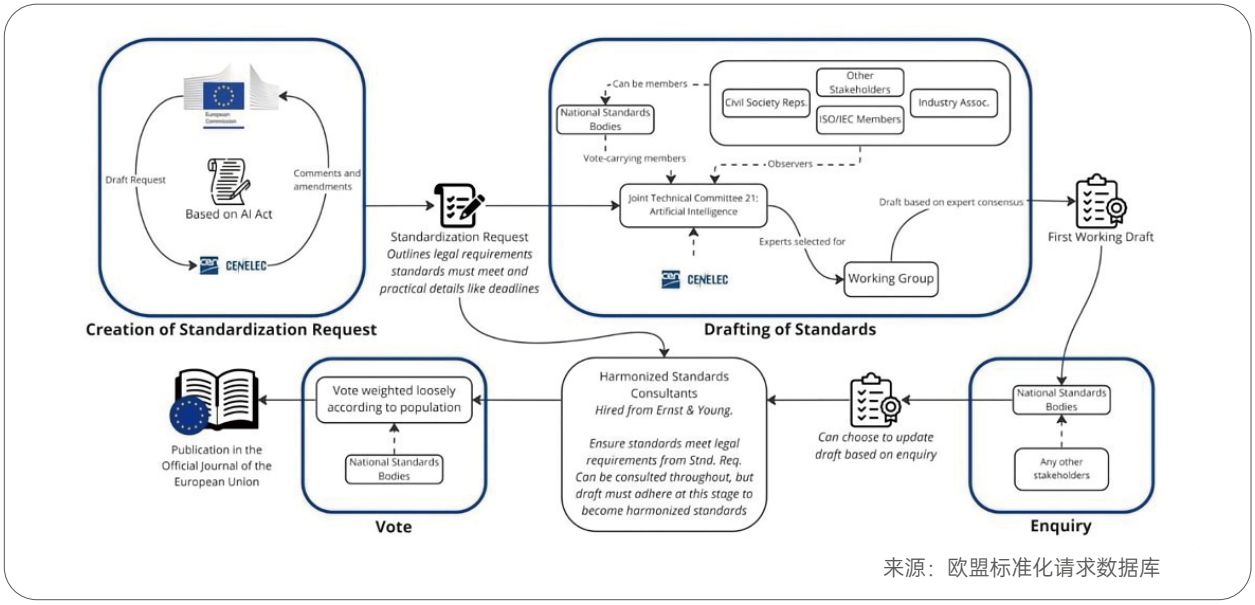
**电气电子工程师协会（IEEE）以伦理治理为核心，逐步构建从标准制定到培训认证的全生命周期标准生态系统。一是确立以其人工智能标准委员会（AISC）为核心的标准化主体。**IEEE作为全球最大的专业技术组织，其下辖的标准协会（IEEE SA）负责制定技术与治理标准，并由其设立AISC统筹人工智能伦理与治理标准研制。此外，IEEE人工智能标准还分布于通信协会（ComSoc）、计算机学会（CIS）、机器人与自动化协会（RAS）等多个技术协会（见图1）。**二是构建以伦理治理为目标的标准体系。**IEEE早期通过《人工智能设计的伦理准则》白皮书，提出伦理优先的系统设计理念；并在此基础上推动P7000系列标准研制，探索将伦理与人类价值观嵌入人工智能系统全流程。该系列涵盖从系统构思，到建模、测试与部署的完整链条，如IEEE 7000-2021《解决系统设计中的伦理问题的建模过程》、IEEE 7003-2024《算法偏见的处理》、IEEE 7010-2020《合乎伦理的人工智能与自主系统的福祉度量标准》等标准。此外，IEEE还面向公众免费开放发布超6个月的P7000系列标准，以推动伦理共识广泛普及。**三是建设跨领域、全栈式的标准生态。**IEEE推动内部人工智能标准资源整合，其标准分布在39个技术协会中，涵盖计算、云、芯片、冷却、网络及行业应用等多个方向。此外，IEEE计划加速内容真实性、大模型治理等标准立项，并推动伦理治理标准制定、CertifAIEd培训与认证的上下游联动，打造伦理驱动与产业落地并重的全球人工智能标准生态系统。


来源：IEEE

图1 IEEE人工智能标准化相关内部组织（部分）

## （二）全球主要经济体标准化行动

**欧盟以三大区域性标准化组织协同推进《人工智能法》协调标准，并通过国际化参与提升标准领域影响力。一是围绕欧盟《人工智能法》开展协调标准工作。**2023年5月，欧盟委员会正式发布标准化需求单（M/593），授权欧洲标准化委员会（CEN）与欧洲电工标准化委员会（CENELEC）合作制定新的欧洲标准或采用国际标准，以支持2024年8月起正式生效的欧盟《人工智能法》[4]（见图2）。该需求单规定了标准制定的10大关键领域，聚焦于风险分级分类中的高风险与有限风险人工智能系统，侧重制定用于验证合规性的可执行方法，突出了法律配套标准的准强制性。**二是对齐国际组织以提升标准制定影响力。**基于维也纳协议、法兰克福协议，以及欧洲电信标准化协会（ETSI）与ITU签订的谅解备忘录等合作文件，CEN、CENELEC与ETSI三大欧洲标准化组织，与ISO、IEC和ITU形成了业务上的对应合作关系，并确立了国际标准化优先原则。同时，通过27个内部主权国家、4个欧洲自由贸易联盟国家以及其他伙伴国家，共计34个国家标准化组织的投票协调一致性，提升了欧盟在各国际标准组织中的影响力。**三是通过欧洲标准化组织统筹协调促进人工智能健康发展。一方面，**CEN与CENELEC第二十一联合技术委员会（CEN/CLC JTC 21）负责统筹制定人工智能标准，聚焦风险管理、数据治理、透明度、稳健性和合规性等领域。截止2025年6月，CEN/CLC JTC 21已发布15项相关标准，其中12项采用了ISO和IEC的标准[5]。**另一方面，**ETSI协调参与CEN和CENELEC的标准工作，重点关注人工智能系统安全。受机构重组的影响，ETSI虽未被指定为标准化需求中的牵头机构，但欧盟相关标准及成果仍需考虑ESTI人工智能行业规范组相关工作，尤其是保护人工智能系统本身以及防止人工智能技术被用于攻击。




来源：欧盟标准化请求数据库

图2 欧盟《人工智能法》协调标准制定示意图

[4] 欧盟标准化请求数据库. M/593号标准化请求[EB/OL]. 2023-05-22. https://ec.europa.eu/growth/tools-databases/enorm/mandate/593_en
[5] CEN. CEN/CLC/JTC21-人工智能已发布标准[EB/OL]. 2025-06-12. https://standards.cencenelec.eu/dyn/www/f?p=205:32:0::::FSP_ORG_ID,FSP_LANG_ID:2916257,25&cs=1827B89DA69577BF3631EE2B6070F207D

**美国以自愿性行业规范、风险导向评估和国际主导战略为特征，推动安全创新并维护全球标准主导地位。一是明确自愿性共识在标准制定中的核心地位。**2024年12月，美国众议院人工智能工作组发布的《两党众议院人工智能工作组报告》明确指出，"美国自愿共识标准方法的最大优势在于其自下而上、基于规则、多利益相关方的过程，在这个过程中，技术优势占上风"[6]。2025年3月，美国国家标准学会（ANSI）在对"人工智能行动计划"的回函中强调，私营部门主导的标准化体系与公私合作关系对于维持美国在人工智能领域的领先地位至关重要，应确保所有利益相关方参与人工智能标准制定工作[7]。**二是发挥美国国家标准与技术研究院（NIST）在标准全局中的领导作用。**美国NIST负责领导国内与国际技术标准制定，以促进人工智能创新和公众信任，其主要任务包括：协调联邦层面标准、参与制定全球标准，以及推动《人工智能风险管理框架》纳入国际标准（见图3）[8]。2025年3月，NIST宣布启动人工智能标准零草案试点，计划加快标准制定进程，融合利益相关方的专业知识与多元视角，以满足人工智能领域标准需求和释放创新活力[9]；7月，NIST发布首份《人工智能测试、评估、验证与确认（TEVV）标准零草案大纲》，旨在与当前及未来ISO和IEC的测试标准相兼容，为相关标准应用提供背景支撑[10]。**三是主要研究机构从"安全监管"转向"标准制定与创新赋能"。**2025年6月，美国商务部将美国人工智能安全研究所（USAISI）调整为人工智能标准与创新中心（CAISI），其职责重点涉及：科学驱动与行业协作的自愿性标准制定、风险导向的人工智能能力评估与标准制定，国际主导与跨部门协调的标准化战略[11]。此举反映美国的人工智能研究机构正从单一的安全监管领域转向推动标准制定与创新赋能，旨在将风险管理融入创新发展中，并通过国际标准领域话语权来提升竞争力。

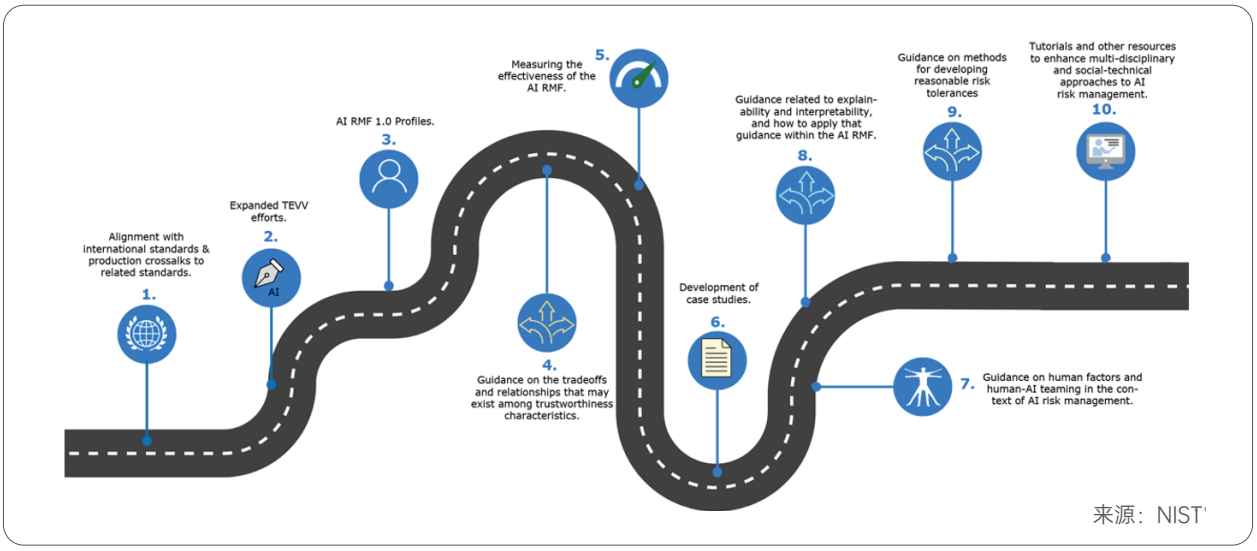



图3 美国《人工智能风险管理框架》路线图

[6] ANSI. 众议院人工智能工作组报告为人工智能政策提供建议，强调标准的作用[EB/OL]. 2024-12-18. https://www.ansi.org/standards-news/all-news/2024/12/12-18-24-house-ai-task-force-report-offers-recommendations-for-ai-policy
[7] ANSI. ANSI代表美国标准化团体向OSTP信息请求提交关于人工智能行动计划的回应[EB/OL]. 2025-03-14. https://www.ansi.org/standards-news/all-news/2025/03/3-14-25-ansi-submits-response-to-ostp-rfi-on-ai-action-pla
[8] NIST. 人工智能风险管理框架（AI RMF 1.0）路线图[EB/OL]. 2023-03-14. https://www.nist.gov/itl/ai-risk-management-framework/roadmap-nist-artificial-intelligence-risk-management-framework-ai
[9] NIST. 人工智能标准"零草案"试点项目旨在加速标准化、拓宽输入[EB/OL]. 2025-03-27. https://www.nist.gov/artificial-intelligence/ai-research/nists-ai-standards-zero-drafts-pilot-project-accelerate
[10] NIST. 欢迎就人工智能测试、评估、验证和确认标准"零草案"大纲发表意见[EB/OL]. 2025-07-29. https://content.govdelivery.com/accounts/USNIST/bulletins/3eb9dd6
[11] 美国商务部. 商务部长霍华德·卢特尼克关于将美国人工智能安全研究所转变为支持创新、支持科学的美国人工智能标准和创新中心的声明[EB/OL]. 2025-06-03. https://www.commerce.gov/news/press-releases/2025/06/statement-us-secretary-commerce-howard-lutnick-transforming-us-ai

**英国将参与全球规则制定作为核心目标，通过资源共享、社区建设和国际合作推动标准化进程。一是国家战略聚焦标准合作定位。**2021年9月，英国《国家人工智能战略》将参与全球标准制订作为重点，强调"从一开始就将准确性、可靠性、安全性等信任要素融入人工智能技术"，明确英国将"在全球人工智能技术标准制定中做出独特贡献"，扩大与志同道合伙伴的国际合作，确保"全球标准由广泛专家按照共同价值观制定"[12]。**二是英国标准协会（BSI）标准化组织推动国际标准落地。**作为ISO创始成员和多个重点标准的牵头单位， BSI深度参与了ISO和IEC的多项人工智能标准制定工作，并大力推动其转化为国内标准（见图4）。2024年1月，BSI率先在国内适用推广BS ISO/IEC 42001:2023，指导组织建立安全、可靠的人工智能管理体系[13]。同时，BSI还在全球推广ISO和IEC制定的其他人工智能标准，为国内外企业和研发团队提供培训、认证、测试与评估服务。**三是英国人工智能标准中心（AI Standards Hub）推动全球交流合作。**2022年10月，在英国政府的支持下，艾伦·图灵研究所、BSI与国家物理实验室牵头成立AI Standards Hub，以共享标准资源、汇聚专家力量并推动国际对话[14]。2025年3月，AI Standards Hub举办的首届人工智能标准峰会邀请了来自不同国家政府、国际标准组织、产业界与学术界的高层代表，共同探讨国际标准化、基础模型治理，以及人工智能安全与标准化社区协作等标准化议题[15]。

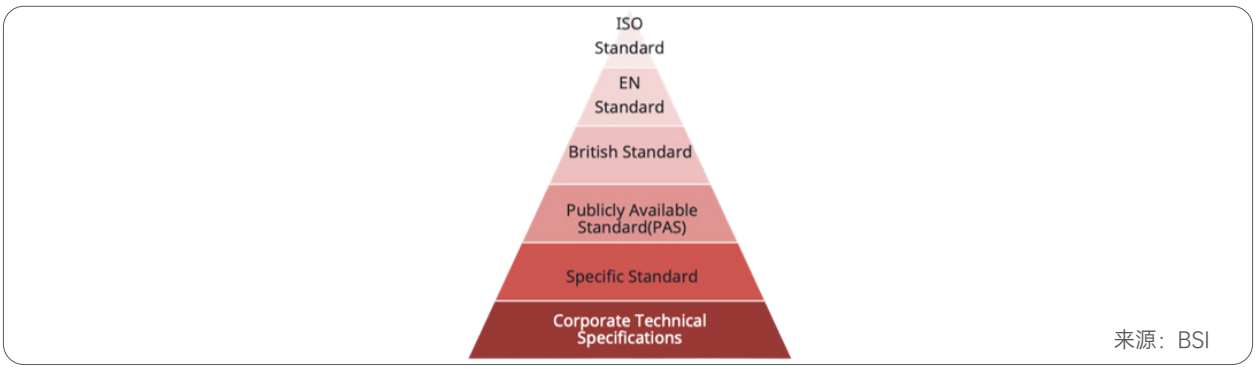
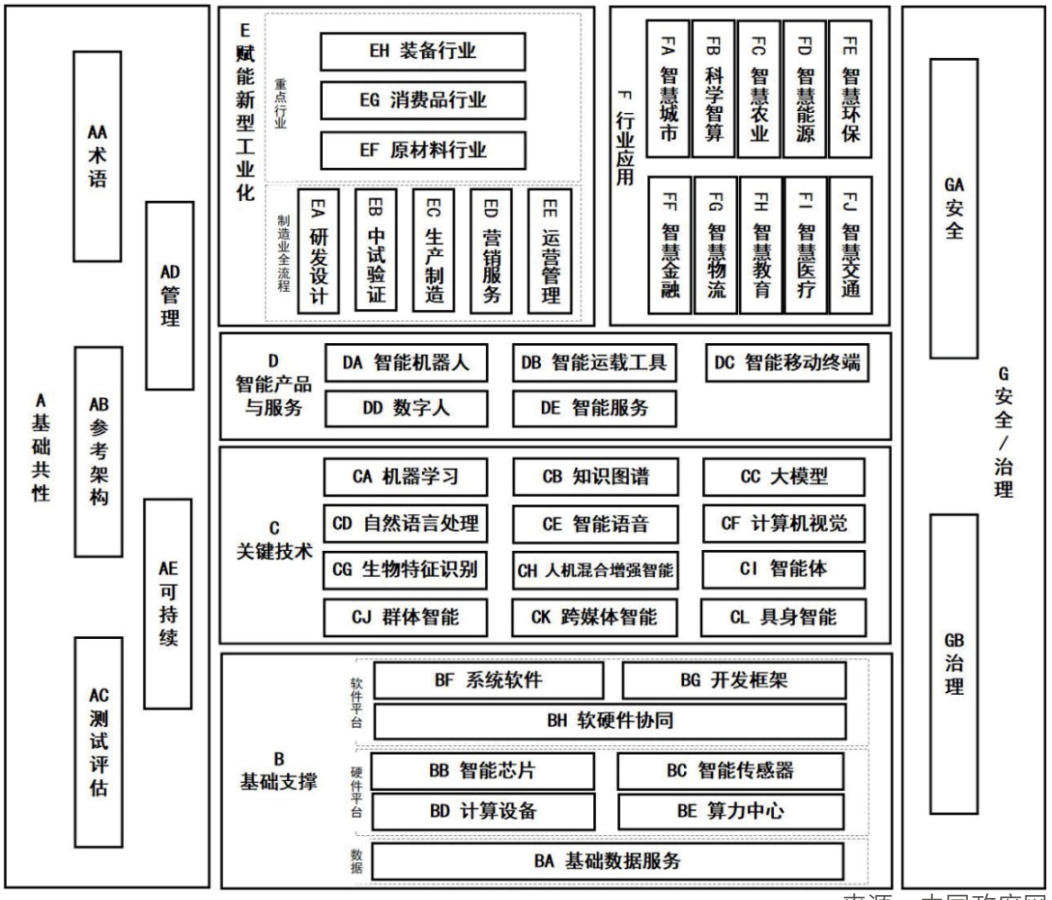



图4 英国BSI人工智能标准协调等级

**中国充分发挥政府主导、产业协同与企业参与优势，呈现出体系布局、多元协同与国际引领特征。一是形成治理倡议与治理框架"走出去"双螺旋布局。**2023年10月，中国发布《全球人工智能治理倡议》（以下简称《倡议》），明确提出应形成具有广泛共识的治理框架与标准规范，推动技术安全、可靠、可控与公平发展，成为助力全球人工智能标准化的中国名片。为全面落实《倡议》，2024年9月和2025年9月分别发布《人工智能安全治理框架》1.0版和2.0版，提出人工智能安全治理框架性方案，并动态调整风险分类，优化完善防治措施，推动人工智能协同共治、普惠共享。**二是推进国内统筹与国际对齐战略部署。**2024年6月，工业和信息化部等四部门联合印发《国家人工智能产业综合标准化体系建设指南（2024版）》，明确到2026年新制定50项以上相关国家/行业标准，并参与制定20项以上国际标准，加速形成支撑人工智能高质量发展的标准体

[12] 英国政府. 国家人工智能战略[EB/OL]. 2022-12-18. https://www.gov.uk/government/publications/national-ai-strategy
[13] BSI. 发布首创的全球指南，以支持负责任的人工智能管理[EB/OL]. 2024-01-16. https://www.bsigroup.com/en-GB/insights-and-media/media-centre/press-releases/2024/january/first-in-kind-global-guidance-to-support-responsible-ai-management-published/#:~:text=16%20January%202024%3A%20A%20first,global%20guidelines%20for%20the%20technology
[14] 艾伦·图灵研究所. 启动人工智能标准中心[EB/OL]. 2022-10-12. https://www.turing.ac.uk/events/launching-ai-standards-hub
[15] 人工智能标准中心. 人工智能标准中心全球峰会[EB/OL]. 2025-03-17. https://aistandardshub.org/global-summit/

系（见图5）[16]。2025年3月，国家网信办等四部门联合发布《人工智能生成合成内容标识办法》，并配套实施《网络安全技术 人工智能生成合成内容标识方法》强制性国家标准及其实践指南，体现了以问题导向、精准施策、动态敏捷的人工智能安全治理理念[17]。2025年8月，国务院印发《关于深入实施"人工智能+"行动的意见》，明确要求重点领域人工智能赋能任务与基础支撑能力建设，其中特别提到加快重点领域人工智能标准研制，推动跨行业、跨领域与国际化标准的联动。**三是推进跨领域联动标准化组织布局。国家标准层面，**全国网络安全标准化技术委员会（TC260）对人工智能安全相关国家标准进行统一技术归口，组织研制《人工智能安全标准体系》，围绕中国人工智能安全治理急需，发布生成式人工智能服务安全、训练数据安全、数据标注安全、生成合成内容标识等国家标准，有序推动人工智能应用安全分类分级、安全能力成熟度评估、涉及未成年人应用等国家标准研制。全国数据标准化技术委员会（TC609）以数据要素流通利用为核心，聚焦数据资源、数据技术与智慧城市建设开展标准化研究。全国信息技术标准化技术委员会人工智能分技术委员会（TC28 SC42）对齐ISO/IEC JTC1 SC42国际标准，重点关注基础软硬件、通用技术、大模型与智能体等关键方向。全国智能技术社会应用与评估基础标准化工作组（SWG35）聚焦生成式人工智能、智能政务等方向，系统构建相关标准体系并组织立项研




来源：中国政府网

图5 中国人工智能标准体系结构图

[16] 中国政府网. 关于印发国家人工智能产业综合标准化体系建设指南（2024版）的通知[EB/OL]. 2024-06-05. https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.htm

[17] 中国政府网. 四部门联合发布《人工智能生成合成内容标识办法》[EB/OL]. 2025-03-15. https://www.gov.cn/lianbo/bumen/202503/content_7014283.htm

制，围绕人工智能社会实验、智能技术政务应用等领域提出国家标准立项建议。**行业标准层面，**中国通信标准化协会（CCSA）依托覆盖全网的信息通信基础，在人工智能、云计算、区块链与6G等方向推进前沿标准制定，形成协会牵头、产业参与、任务驱动的协作机制。工业和信息化部人工智能标准化技术委员会（MIIT/TC1）统筹协调工业和信息化领域人工智能行业标准，聚焦基础共性、关键基础技术、产品服务、赋能应用、安全治理五大领域行业标准体系，形成由8个工作组组成的专业网络，成员单位超500家。

**全球南方国家在标准化中的代表性不足，但其人口规模与发展诉求决定其在全球标准秩序重塑中不可或缺。**金砖国家、东南亚国家联盟（简称"东盟"）与非洲联盟（简称"非盟"）等区域组织，正依托各自政治经济联盟，探索包容性人工智能标准化路径，努力推动全球规则体系，增强南方声音。**一方面，依托区域合作机制强化集体发声。**三大组织均强调发展中国家在全球治理中的主体地位，主张发展权、技术公平与主权尊重，旨在打破标准制定中的制度性壁垒。2025年7月，金砖国家领导人签署《金砖国家领导人关于人工智能全球治理的宣言》，提出治理工作应以国家监管框架和《联合国宪章》为准则，呼吁建立联合国主导、南方国家广泛参与的全球治理框架[18]；2025年3月，东盟发布《负责任人工智能路线图（2025-2030）》，提出以技能与能力建设、公平与包容、治理与参与、整合与合作为区域人工智能发展的四大支柱[19]；2024年8月，非盟在《非洲大陆人工智能战略》中强调，技术发展需服务于本地优先事项，以推动包容性人工智能治理，助力非洲繁荣[20]。**另一方面，以联合研究与制度文件凝聚标准共识。**南方国家虽尚未建立系统性人工智能标准体系，但普遍采用顶层战略引领、多边协作推进的路径，为未来标准制定奠定制度基础。金砖国家通过设立人工智能研究小组，聚焦统一治理框架与技术伦理问题，推动成员间开展前期标准协调研究；东盟结合《人工智能伦理治理指南》《数字经济框架协议》等规范，推动法律法规与技术标准协同接轨，同时推动跨国能力建设与场景试点；非盟在战略指引下，推动各成员国建立本地人工智能治理框架，并在农业、健康与教育等关键领域推动人工智能技术落地与标准适配。此外，中东国家通过国家战略与开放模型推动标准化能力提升，如阿联酋Falcon系列模型在推动人工智能技术可及性、透明性与区域标准协同上发挥示范作用[21]；巴西、智利与哥伦比亚等拉丁美洲国家借鉴欧盟风险分级方法推动立法[22]，并通过《泛美数据与人工智能治理框架》等区域文件为未来标准化工作奠定合作基础[23]。

[18] 金砖国家. 金砖国家领导人宣言谴责战争 呼吁全球治理改革[EB/OL]. 2025-07-17. https://brics.br/en/news/collabs/collaborative-communication/brics-leaders-declaration-condemns-wars-and-calls-for-reform-of-global-governance
[19] 东盟. 东盟负责任人工智能路线图（2025-2030年）[EB/OL]. 2025-02-28. https://asean.org/book/asean-responsible-ai-roadmap-2025-2030/
[20] 非盟. 非洲大陆人工智能战略[EB/OL]. 2024-08-09. https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy
[21] 时代周刊. 阿联酋的使命是成为人工智能强国[EB/OL]. 2024-03-22. https://time.com/6958369/artificial-intelligence-united-arab-emirates/
[22] White & Case. 促进创新还是降低风险？拉丁美洲的人工智能监管[EB/OL]. 2024-11-18. https://www.whitecase.com/insight-our-thinking/latin-america-focus-2024-ai-regulation
[23] 美洲国家组织. 美洲数据治理和人工智能框架（MIGDIA）[EB/OL]. 2024. https://www.oas.org/ext/en/democracy/inter-american-framework-on-data-governance-and-artificial-intelligence-migdia

# 02 全球人工智能标准发展态势

## （一）全球人工智能标准发展加快提速

**从标准研究内容上看，全球人工智能标准朝着体系化方向发展。一方面，形成以技术、应用与治理为分层的标准体系。美国**以底层技术规范为基础、辅以治理方法工具为支撑，在NIST主导下提出涵盖术语、数据、人机交互、测试评估、风险管理、安全与可信度等九大重点领域[24]，将TEVV方法、风险导向管理、透明度工具和隐私安全等作为优先方向[25]；**英国**将伦理治理与应用导向相结合，其在实践中突出治理、伦理与安全优先，将算法偏见、数据隐私、透明度与伦理道德作为关注焦点[26]，同时把人工智能列为五大关键技术之一[27]，着重推动其在工业、金融、医疗、教育、交通等行业的应用规范；**中国**全面覆盖从底层技术、产业应用到安全治理的标准全链条，在《国家人工智能产业综合标准化体系建设指南（2024版）》中提出涵盖基础共性、基础支撑、关键技术、智能产品与服务、赋能新型工业化、行业应用以及安全/治理7个部分组成的体系框架。**另一方面，研究重心从基础技术向应用赋能与安全治理延伸。在应用赋能方面，**对推动创新与产业赋能的标准需求显著增长，特别是在智能产品与行业应用方面，标准正为新兴场景提供明确规范。例如，ITU-T F.748.46《通用智能体能力要求及评估方法》对智能体在复杂任务的执行过程中提出了能力要求和评估框架；ITU-T Y.4472《智能城市中人工智能和物联网的协同数据管理》则为智慧城市的跨设备数据协同提供了标准化路径。**在安全治理方面，**伴随人工智能能力提升而来的技术性与社会性风险，也推动安全与治理类标准迅速兴起。ISO/IEC 42001:2023作为全球首个人工智能管理标准，为组织人工智能治理提供了制度化框架；ISO/IEC 42005:2025要求组织全面评估其对个人、社会和环境的潜在影响，有力推动责任落实与信任构建。

**从标准制定过程上看，呈现出技术敏捷性与治理复杂性并存的特点。一方面，前沿技术类标**

[24] NIST. 美国在人工智能领域的领导地位：关于联邦机构参与制定技术标准及相关工具计划[EB/OL]. 2019-08-09. https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf
[25] NIST. 全球人工智能标准合作计划[EB/OL]. 2024-07. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-5.pdf
[26] 国际知识产权法学协会. 英国上议院呼吁知识产权部长解决人工智能监管问题[EB/OL]. 2023-07. https://iipla.org/uks-lords-call-on-ip-minister-to-address-ai-regulation/
[27] 英国政府. 启动使英国成为国际科技超级大国的计划[EB/OL]. 2023-03-22. https://www.gov.uk/government/news/plans-to-make-uk-an-international-technology-superpower-launched

**准制定流程不断优化。**主要国际标准化组织普遍采用快速立项与加速评审机制，以适应技术迭代周期，如ITU在前沿技术类标准审批流程中引入在线快速审批模式，ISO和IEC则全面采用模块化、可组合的在线制定模式加快标准落地。此外，美国NIST的零草案试点也有望为TEVV标准快速落地提供制度保障。**另一方面，社会治理类标准进展相对缓慢。**人工智能技术应用发展迅速，相应地，人工智能技术性能类标准制定较快，而人工智能安全治理配套标准相对缓慢。与此同时，各主要经济体不同治理路径需配套不同标准，相应地推动安全治理国际标准达成共识则需要更长时间在各国间进行磋商。**此外，**全球南方国家在标准制定中的结构性缺位进一步加剧了治理标准推进的缓慢。尽管国际标准组织普遍强调多元包容性，但资源匮乏、技术壁垒和话语权不足等主客观因素使其在核心工作组中的影响力有限。例如，东盟、非盟国家的优先事项集中在健康、教育、粮食安全等基本需求，而对人工智能治理与伦理标准普遍持有更加审慎的态度。

**从标准应用程度上看，国内外标准在不同国家和地区的标准体系和法规实施中呈现多样性。一是南方国家适应性采用需求突出。**全球南方国家受限于资源与能力匮乏，为加快人工智能技术引进与产业部署，普遍倾向引入国际标准或其他经济体的成熟标准。**二是国际标准的主要参与国家积极推动国内转化。**深度参与国际标准制定的主要经济体，倾向于将其标准直接等采为国内或区域标准，从而在法规配套、产业推广和技术生态建设中保持一致性。例如，英国BSI深度参与ISO/IEC 42001:2023编制工作，积极推动国内标准转化并大力推广全球标准认证与培训；欧盟及其伙伴国家的34个标准化机构在ISO、IEC、ITU等组织中享有广泛投票权，CEN和CENELEC在制定区域标准时倾向直接或间接采用相关国际标准。**三是国家和区域内部准法律效力不断增强。**在标准化体系较为完善的国家和地区，技术标准与法律法规紧密绑定，形成具有法律效力的实施机制。中国《网络安全技术 人工智能生成合成内容标识方法》强制性配套标准向上衔接《人工智能生成合成内容标识办法》，向下发布6项网络安全标准实践指南，明确不同类型内容的标识要求，旨在防范内容风险；2025年7月，欧盟发布《通用人工智能实践准则》"类标准化文件"，为《人工智能法》中通用人工智能要求的落地提供技术实施指南，并明确将其作为减轻企业负担和提高法律确定性的重要参考文件[28]。

## （二）负责任的人工智能标准研究成为全球焦点

**负责任的人工智能标准化已从理念共识进入可操作规范的加速落地阶段。**联合国教科文组织（UNESCO）于2021年11月发布的《人工智能伦理问题建议书》，明确所提出的建议是通过全球方法制定的标准工具，以期指导人工智能技术负责任的发展。ITU借助人工智能向善全球峰会，连续多年举办"负责任的人工智能"系列会议，推动多边协作与跨界对话，围绕伦理、安全、公平与透明等原则，构建负责任人工智能的治理愿景与实践路径。2023年11月，世界互联网大会发布《发展负责任的生成式人工智能研究报告及共识文件》，结合全球实践案例，提出十条共识，引

[28] 欧盟委员会. 欧盟关于通用人工智能模型的规则开始适用，带来更多的透明度、安全性和问责制[EB/OL]. 2025-08-01. https://digital-strategy.ec.europa.eu/en/news/eu-rules-general-purpose-ai-models-start-apply-bringing-more-transparency-safety-and-accountability

[29] 世界互联网大会. 发展负责任的生成式人工智能研究报告及共识文件[EB/OL]. 2023-11-09. https://cn.wicinternet.org/2023-11/09/content_36952741.htm

起国际社会广泛关注 。基于此，本报告提出涵盖**可持续发展、能力建设、普惠创新、风险管理、安全可靠、数据治理与隐私保护等**方面的标准化趋势，进一步探讨国际标准组织、主要经济体与产业界负责任的人工智能标准化行动。

**国际标准组织重点推动形成全球安全、可信与合乎伦理的负责任治理理念。一是通过国际合作平台承诺安全可控。**2023年4月，ISO、IEC和ITU在世界标准合作组织（WSC）框架下联合回应未来生命研究所（FLI）对于开发高级人工智能系统的担忧，强调通过标准化过程保障人类监督与问责，从而支撑法规框架并提升用户信心；同时，他们还呼吁共同制定基于共识的国际标准并推动其采用[30]。**二是ITU以智能向善为目标推动公众信任。**ITU通过人工智能向善（AI for Good）倡议积极响应联合国可持续发展目标，一方面持续推进"人工智能准备度（AI Readiness）"进程，明确互操作性与安全标准是增强公众信任、确保不同供应商与用户均可接受解决方案的关键[31]；另一方面，系统汇集ISO、IEC、ITU、IEEE和国际互联网工程任务组（IETF）等国际组织标准化成果，建设全球可信赖、负责任与合乎道德的人工智能标准数据库，为企业、政策制定者和监管者提供权威参考[32]。**三是ISO/IEC构筑可信赖体系完善全生命周期治理。**ISO和IEC在现有覆盖人工智能生态系统的标准基础上，进一步聚焦生成式人工智能数据、模型与应用领域，提出基于自保安全（Security）与无害安全（Safety）等语境，持续构建安全、可信赖的人工智能标准体系。**四是IEEE以伦理治理为导向促进跨界生态协同。**IEEE聚焦伦理导向下的人工智能系统设计，持续落地自主智能系统伦理全球倡议，强调推动技术与道德考量的深度融合，并通过开源共享与跨界生态促进全球人工智能治理与标准的协同发展。

**主要国家和地区在将负责任特征转化为标准体系构建核心驱动力。一是欧盟打造以风险管理为核心的全球示范区。**CEN和CENELEC在欧盟《人工智能法》框架下，将可信、安全、以人为本等法律原则转化为可落地技术细节，依托标准化请求制定覆盖高风险人工智能系统全生命周期的协调标准。同时，其通过与ISO和IEC等国际组织对接，保持与NIST和经济合作与发展组织（OECD）等机构对话，降低技术壁垒并提升欧盟标准国际采用程度，从而在全球人工智能治理中发挥示范效应。**二是美国探索负责任创新与行业自治融合路径。**NIST在人工智能风险管理框架中明确可信系统的有效与可信赖、安全、可靠与弹性、负责与透明度、可理解与可解释性等七大要素，并通过《人工智能全球标准参与计划》中提出的可信人工智能研究、国际标准映射与联合测试方法推动国内行业自治标准上升为国际标准。同时，NIST在TEVV领域的人工智能标准零草案试点，有助于缩短标准出台时间并协调多方利益相关者，在强化责任治理的同时又保留了行业创新的灵活性。**三是中国全面推动安全治理与普惠发展相结合。**2025年7月，MIIT/TC1在《国家人工智能产业综合标准化体系建设指南（2024版）》的"安全/治理"基础上，发布《工业和信息化领域人工智能安全治理标准体系建设指南（2025版）》进一步细化人工智能安全治理标准体系

[30] WSC. IEC、ISO和ITU回应FLI公开信：国际标准有助于确保安全和负责任的人工智能开发[EB/OL]. 2023-04-06. https://www.worldstandardscooperation.org/ai-and-standards/
[31] ITU. 人工智能准备度：面向标准化准备程度框架的分析[EB/OL]. 2024-09-01. https://www.itu.int/hub/publication/t-ai4g-ai4good-2024-2/
[32] ITU. 人工智能标准交流数据库欢迎投稿[EB/OL]. 2025-07-21. https://www.itu.int/hub/2025/07/ai-standards-exchange-database-welcomes-contributions/#:~:text=A%20new%20AI%20Standards%20Exchange,innovations%20to%20achieve%20global%20impac

结构，为推动中国人工智能产业高质量发展提供坚实技术支撑。此外，国内标准化组织结合智慧城市、医疗健康、公共服务等领域的应用实践经验，推动制定普惠发展标准规范，既加快国内数字化转型，也为全球负责任人工智能发展贡献中国方案。

**产业界正由合规执行者加速转型为规则塑造者，将负责任实践贯穿于人工智能全生命周期。一是主动采纳标准化规范强化合规实践。**企业不仅满足法律要求，还积极签署和采纳诸如欧盟《通用人工智能实践准则》等更高规格的行业自律文件，通过对标协调标准提前适配法规要求、提升技术与治理水平。**二是深度参与国内外标准组织体系化建设。**谷歌、微软、OpenAI、阿里巴巴、华为、科大讯飞等中外企业派遣专家常年活跃于ITU、ISO、IEC、IEEE等国际组织，以及本国或地区标准化组织，围绕数据质量、算法公平性、可解释性、风险评估、隐私保护和伦理治理等核心议题贡献经验智慧。**三是积极参与标准化工作，推动标准应用实践。**企业主体不仅参与讨论，还牵头制定涵盖大模型、智能计算、人工智能管理体系、数字人等领域国际标准，并通过跨国合作和产业联盟推动标准在全球场景落地。例如，英伟达、安谋与英特尔联合发布FP8低精度浮点开放交换格式标准，促进不同硬件与软件平台之间的互操作性[33]；谷歌牵头发起A2A（Agent-to-Agent）智能体通信协议，联合五十余家合作伙伴构建开放、安全、高效的智能体协作网络，推动形成"智能体互联网"时代的基础性开放标准[34]。**四是通过开源社区推动形成标准化生态。**LangChain、Hugging Face、MLCommons等开放框架与社区，通过共享模型接口、评测基准和工具集，建立了从模型训练到评估验证的开放技术规范，加速行业共识凝聚与标准下沉应用。总的来说，越来越多的企业在可持续发展、能力建设、普惠创新、风险管理、安全可靠、数据治理与隐私保护等方面践行负责任的实践，有力推动全球人工智能标准化发展（详见附录）。

## （三）国际间人工智能标准互操作性势在必行

**国际社会对人工智能标准互操作性的需求日益迫切，涵盖技术、经济、科研与社会服务等多个层面。技术层面，互操作性旨在打破不同系统、模型和平台之间的壁垒。**通过统一的数据格式、元数据规范和API接口标准，可让在单一平台训练的模型直接调用其他平台的数据或迁移部署到不同环境，从而提升人工智能资产的流动性。英特尔在人工智能芯片中引入互操作性设计后，减少了重复集成和适配工作，带来超过110万美元的效率提升[35]。**经济层面，互操作性是降低贸易壁垒和推动全球一体化的重要手段。**基于国际共识标准的产品认证可在多国直接互认，减少本地化适配和重复认证成本，从而加快市场准入速度。尤其在云计算领域，当用户可以在亚马逊云（AWS）、微软云（Azure）、阿里云等不同云平台之间部署跨云人工智能应用时，能够有效促进市场的竞争和创新[36]。**科研层面，互操作性成为加速全球协同创新的重要催化剂。**统一的标准能打通科研数据和模型的"孤岛"，让跨国团队高效共享和整合资源。以现有国际合作为例，由多

[33] 英伟达. 英伟达、安谋和英特尔发布FP8标准化规范作为人工智能的交换格式[EB/OL]. 2022-09-14. https://developer.nvidia.com/zh-cn/blog/nvidia-arm-and-intel-publish-fp8-specification-for-standardization-as-an-interchange-format-for-ai/
[34] A2ACN. A2A Protocol技术文档[EB/OL]. https://agent2agent.info/zh-cn/docs/
[35] 英特尔. 执行摘要：英特尔人工智能的总体经济影响™[EB/OL]. 2021-08-19. https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/Executive-Summary-The-Total-Economic-Impact-of-Intel-AI/post/1335739
[36] 欧洲监管中心. 云计算服务的竞争与监管及欧盟政策综述[EB/OL]. 2024-04. https://cerre.eu/wp-content/uploads/2024/02/REPORT.CERRE_.FEB24.CLOUDS.pdf

国研究机构共同推动的材料设计开放数据库集成（OPTIMADE）联盟，建立了标准化数据接口，使全球材料数据库互联互通，从而显著加快了新材料发现的进程。**社会服务层面，互操作性是提升全球性危机应对能力的关键基础。**采用统一数据和接口交换标准可在公共卫生事件或自然灾害中实现跨国系统的高效协同与资源调配。在地震、洪水、飓风等灾害发生后，来自不同国家和组织的救援力量需要协同作战，而高效、互通的人工智能辅助系统可将灾难紧急响应时间减少58%、幸存率提高32%[37]。

**与此同时，全球多方通过组织合作、跨国互认与政策调整推进互操作性标准化实践。一是国际标准组织通过联合行动应对标准碎片化。**近年来，ITU、ISO和IEC依托世界标准合作（WSC）框架，重点加强内容来源、信任与真实性、水印、资产标识符，以及权利声明等领域标准互操作性，旨在保护个人信息权益和培养对数字生态系统的信任[38]。2024年10月，三方在ITU世界电信标准化全会（WTSA）期间举办首届国际人工智能标准峰会；2025年12月，将继续举办首尔国际人工智能标准峰会，意在进一步凝聚跨国标准合作共识。**二是国家间采取双边与多边机制推动标准互认。**美国与欧盟在贸易与技术委员会（TTC）下设立关键术语、标准工具与监测评估三个工作组，重点推进《可信人工智能和风险管理联合路线图》的实施，为标准互操作奠定技术与制度基础[39]。中国-东盟数据安全倡议提出技术互认、标准互鉴和监管互助为核心的三阶信任模型，推动区域内47项共通标准制定，强化数字经济互联互通[40]。新加坡与美国在关键及新兴技术对话（CET）中完成AI Verify框架与NIST《人工智能风险管理框架》的映射，实现治理框架的互操作[41]。**三是多国与区域组织通过立法与政策调整对接国际标准。**中国修订《采用国际标准管理办法》，明确ISO、IEC和ITU发布的国际标准可等同或转化为国家标准，以加快国内外标准的一致性。欧盟在《人工智能法》配套标准化请求中，要求CEN和CENELEC参考国际标准成果，以保持法规执行的国际兼容性。新加坡大力推动与ISO/IEC 42001:2023国际标准对齐，从而提升其在国际社会的可接受度与互操作能力。

[37] Pharaoh Soft. 2025年人工智能驱动的灾难响应协调：应急管理的未来[EB/OL]. 2025-04-19. https://pharaohsoft.com/ai-driven-disaster-response-coordination-in-2025-future-of-emergency-management/
[38] WSC. 人工智能和多媒体真实性标准合作：由世界标准合作组织牵头的多利益攸关方倡议[EB/OL]. 2025-07-11. https://www.worldstandardscooperation.org/what-we-do/amas/
[39] 美国贸易代表办公室. 美国-欧盟贸易和技术理事会联合声明[EB/OL]. 2023-05-31. https://ustr.gov/about-us/policy-offices/press-office/press-releases/2023/may/us-eu-joint-statement-trade-and-technology-council#:~:text=The%20groups%20have%20,of%20existing%20and%20emergent%20risks
[40] 中国社会科学网. 中国-东盟加强数字治理合作的有效路径[EB/OL]. 2025-06-27. https://cssn.cn/zkzg/zkzg_gxzkdzyx/zkzg_gxdzyx_yc/202506/t20250627_5883790.shtml
[41] 新加坡数码发展及新闻部. 新加坡和美国将深化人工智能领域的合作[EB/OL]. 2023-10-13. https://www.mddi.gov.sg/newsroom/singapore-and-the-us-to-deepen-cooperation-in-ai

# 03 全球人工智能标准化的挑战和困难

## （一）技术迭代提速与标准建设缺口并存

**人工智能技术迭代速度显著快于标准化进程，使得标准制定长期处于被动跟随状态。**这种缺口可能导致标准在落地时难以完全适配最新技术能力与风险场景，从而影响产业投入效率、安全保障水平和跨领域应用的稳定性。**一是技术爆发式创新与标准化周期存在结构性矛盾。**2025年3月，根据模型评估与威胁研究实验室（METR）提出的"衡量人工智能完成长时间任务的能力"评测结果，前沿人工智能系统完成任务的能力大约每7个月翻一番，远超传统摩尔定律的两年周期[42]。相比之下，典型国际标准化流程往往涉及数十个成员国与多轮投票协商，标准从立项到发布需耗时2-4年，这种以年为单位的标准化节奏难以满足以月为单位的前沿技术更迭需求。此外，即便在研究阶段适配最新技术，标准发布后也会因架构演进和性能跃迁而降低适应性，尤其是在医疗影像诊断、自动驾驶等高风险应用领域。**二是新兴技术场景带来关键标准空白。**以生成式人工智能为例，其大规模普及暴露了深度伪造检测、多媒体真实性验证、大模型可解释性等领域标准缺口。**在深度伪造检测方面，**不同国家和厂商在检测精度、验证流程及数据集规范上缺乏统一要求，导致检测结果难以互认；**在多媒体真实性验证方面，**内容来源和真实性联盟（C2PA）等行业组织积极推动元数据溯源方案，但全球国家和地区的内容治理在跨平台兼容性、地域适配性等方面仍需持续完善；**在大模型可解释性方面，**部分行业虽已提出建立相应的评估指标体系，但其量化方式、验证机制及通用化路径仍缺乏全球共识。**三是行业复杂性加剧标准制定难度。**不同行业在人工智能技术与应用性能、安全及合规等方面存在显著差异，这对标准的适配性提出更高要求。例如，**医疗领域**需严格符合医疗器械监管和患者隐私保护规范，重点考察临床验证、风险管理和伦理合规等方面；**能源领域**需处理大规模实时数据并与关键基础设施安全要求对接，着重考

[42] METR. 衡量人工智能完成长时间任务的能力[EB/OL]. 2025-03-19. https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/

虑系统安全性、稳定性与响应时间要求。此外，医疗、新能源等领域受关注度高，同一行业的不同场景所面临的风险程度不同，不能用行业来定义风险高低。跨行业协调不仅要协调差异性，还要处理不同技术路线的兼容性问题，使得标准制定周期进一步延长。

## （二）产业链条复杂加剧标准制定难度

**人工智能产业链条结构高度复杂、参与主体多元，导致标准制定在利益协调与技术统一上面临显著障碍。**这种复杂性延长了标准研制周期，削弱了标准的通用性与可操作性，也降低了其对产业创新与协同发展的支撑作用。**一方面，产业链高度复杂与技术栈强耦合。**人工智能产业链涵盖算力、算法和数据基础层、深度学习等框架层、大语言模型等技术产品模型层及行业场景应用层四大核心层级，形成上下游高度依赖的技术栈。例如，常见的大模型训练不仅高度依赖硬件GPU算力架构和PyTorch算法框架，还需与应用层的B2B、B2C和B2G业务逻辑适配，标准修改可能影响多个环节的稳定运行。此外，在医疗、金融、物流等不同场景中，需要结合专业需求设计垂类大模型，这导致许多量身定制的行业标准难以上升为通用标准。**另一方面，多元主体利益冲突明显。**不同层级的产业主体在标准化中的立场诉求差异显著，这种分化会导致在标准研制中不易达成共识。个别硬件和云平台提供商倾向于维护封闭生态，降低跨平台标准化的积极性；算法模型研发企业在透明度与商业机密之间权衡，减少算法细节披露；平台工具厂商注重本生态的稳定性，弱化兼容性标准的制定；行业应用的标准需求高度差异化，如医疗领域重安全与隐私保护，金融领域重偏见控制，物流领域重效率优化等；产品服务终端用户则更关注标准的落地可行性与部署成本，采纳意愿不一。

## （三）主要经济体治理理念的差异提高标准共识的门槛

**主要经济体在人工智能治理理念上存在差异，直接影响标准化行动上的路径选择。**这种差异性不仅使国际协商更难达成统一框架，也削弱了现有标准在跨国层面的适用性与互认性。**一方面，不同治理理念导致标准化路径分化。**欧盟坚持以风险治理为核心的法律规制，美国强调创新驱动发展的产业自律，英国主张原则导向治理与跨国协同合作，中国呼吁统筹安全治理与普惠发展。不同治理理念下的路径制定的相关标准存在路径差异，可能影响标准规定实质内容和实施效果上存在偏差。例如，在风险管理领域，欧盟《人工智能法》要求配套的协调标准为高风险系统提供合规路径，而美国则主要依托NIST《人工智能风险管理框架》等指导性文件推动产业自律，两种模式在国际协商中较难以收敛为统一的标准要求。**另一方面，国际标准与国家治理理念偏移加剧标准共识阻力。**在国际人工智能治理标准框架下，涉及风险评估、安全合规与影响评估等重点领域国际标准，可能与部分国家法律法规与监管目标发生冲突，因此相关国家在采纳时会选择本地化改造、延后采用甚至不采纳，从而削弱国际标准的一致性和外溢效应。2025年7月，

CEN/CLC JTC 21主席在世界人工智能大会上表示，欧盟已采用了数十项ISO和IEC制定的国际标准，并参考了ISO/IEC 42001:2023及风险影响评估方法。同时，为更好地推动《人工智能法》落地，欧盟仍将在现有国际标准框架基础上，持续探索和发展更符合区域法律特点的协调标准[43]。

## （四）南方国家现实差距导致标准参与欠缺

**全球南方国家在标准化进程中参与度不足，核心原因在于基础层匮乏、技术人才缺口与代表性不足的多重制约。**这种缺位削弱了国际标准的普惠性与公平性，难以充分反映不同发展阶段与社会土壤下的多元化需求。**一是基础设施薄弱与高质量数据缺失。**许多南方国家在算力设施和数据资源方面存在长期短板，限制了其在标准制定中的技术发言权。**高性能计算方面，**根据数据中心地图（Data Center Map）统计，截止2025年8月，美国共拥有3955个数据中心，而中国（362个）、印度（268个）、巴西（188个）、非洲大陆（223个）等国家和地区的数量远远滞后，这种算力瓶颈使其难以开展大规模模型训练与验证[44]。**高质量数据方面，**印度有22种主要语言，但除了英语之外的印地语、泰米尔语等本土模型开发滞后，2025年政府主导的巴希尼（Bhashini）国家语言翻译计划仍高度依赖英语-本地语言双语数据，缺乏纯本地语言基础模型[45]。**二是核心技术依赖与高端人才短缺并存。核心技术方面，**南方国家的GPU芯片高度依赖英伟达、高通等外国企业，核心算法框架仍以TensorFlow、PyTorch等国外开源工具为主，高精度传感器、人工智能专用芯片等环节尚未形成规模化生产能力，底层技术生态易受国际供应链波动影响。**高端人才方面，**全球南方普遍面临教育体系滞后与人才流失的双重困境：非洲、拉美及部分亚洲国家的科学、技术、工程和数学（STEM）教育资源严重匮乏，农村学校缺乏实验室、计算机和互联网设施；同时，美国又对南方国家形成显著的人才"虹吸效应"，削弱了其持续参与国际标准化的专业能力储备[46]。**三是决策代表性与话语权占比不足。**南方国家在国际标准化议程和决策机制中占比较少，导致其在标准制定中经常处于"接受者"而非"制定者"的被动地位。以全球人工智能伙伴关系（GPAI）为例，其22个创始成员中全球南方仅占2席，会议中南方代表发言时长不足20%，制度正当性明显不足[47]。联合国贸易和发展会议（UNCTAD）数据进一步表明，全球共有118个国家缺席人工智能治理重点议题讨论，其中大部分为南方国家，这造成政策议程与标准内容难以覆盖南方国家的现实需求[48]。

[43] 欧盟驻华标准化SESEC. CEN-CENELEC JTC 21主席Sebastian Hallensleben博士在WAIC 2025发表演讲[EB/OL]. 2025-08-27. https://mp.weixin.qq.com/s/3Um8_-ad5b19xYt3KWtkGw
[44] 数据中心地图. 数据中心[EB/OL]. https://www.datacentermap.com/datacenters/
[45] BBC. 如何让人工智能以22种语言工作[EB/OL]. 2025-08-12. https://www.bbc.com/news/articles/cn0qqzz1e4zo
[46] 保尔森基金会. 全球人工智能人才追踪器2.0[EB/OL]. https://archivemacropolo.org/interactive/digital-projects/the-global-ai-talent-tracker/
[47] 光明网理论. 推动形成多元共治的全球人工智能治理格局[EB/OL]. 2025-08-02. https://theory.gmw.cn/2025-08/02/content_38194147.htm
[48] UNCTAD. 人工智能价值4.8万亿美元的未来：联合国贸易和发展部对分歧发出警报，敦促采取行动[EB/OL]. 2025-04-07. https://unctad.org/news/ais-48-trillion-future-un-trade-and-development-alerts-divides-urges-action

# 04 发展展望与建议

促进全球人工智能标准发展并非单一主体可以独立完成的任务，是一项需要全球协作的系统性工程。为构建一个包容、互通、可持续且有利于创新的全球人工智能标准生态，国际社会需加强协同治理，弥合组织、地域和领域的壁垒，建立高效、透明的多方协作机制；促进包容性发展，促进标准化进程能广泛吸纳不同发展水平经济体的需求，促进全球公平参与，弥合智能鸿沟。

## （一）国际组织着力标准协调，发挥基础引领作用

**国际标准组织**应发挥桥梁作用，增强全球标准制定的协调性、包容性、基础性。**一是**建立更紧密的协同机制。主要国际标准组织应牵头设立常态化的标准协调会议，定期沟通工作计划，识别重叠领域，共同规划跨组织联合项目，从源头上减少标准冲突与重复。**二是**增强发展中国家的实质性参与能力。设立能力建设专项，通过提供会议费用补贴、开展技术培训、推广多语言文档等方式，显著降低发展中国家专家的参与门槛，确保标准的全球代表性与适用性。**三是**推进负责任的标准建设，加强标准前瞻布局。汇聚全球专家共识，继续增强人工智能安全、可信、负责任等基础共性标准制定，并为未来技术趋势预留接口。

## （二）政府部门加强统筹规划，促进标准互通互鉴

**各国政府**是推动标准互鉴、促进技术普惠可及的关键力量，应加快标准化建设统筹布局、多元协作、国内国际接轨。**一是**体系化完善国家人工智能标准发展路线图。各国政府应基于本国产业发展水平和战略目标，发布清晰的国家标准发展战略，明确重点支持领域，引导国内资源与国际标准议程精准对接。**二是**鼓励标准需求探索路径。在具体技术与场景层面，支持区域性、行业性或团体标准先行先试，待实践成熟并形成广泛共识后，再稳步推动其被更广泛采纳，以兼顾创新活力与规范有序。**三是**推动跨国、跨区域标准协调与合作。积极构建双边或多边科研合作与信息共享平台，通过设立跨国专家工作组，就数据治理、隐私保护、安全可控等关键领域开展联合

研究。**四是**加大对国际标准化活动的政策与资源支持。通过设立专项基金，鼓励和支持本国企业、科研机构专家深入参与国际标准组织工作，对做出实质性贡献的机构与个人予以认定和奖励。

## （三）产业界聚集技术贡献与产业协同，加速标准应用转化

**产业界**是技术创新的源泉和标准应用的主体，应主动将最佳实践转化为行业规范。**一是**主动贡献技术方案与最佳实践。成熟企业应积极将成熟的技术方案和行业解决方案贡献给标准制定过程，加速标准的技术验证与产业采纳。**二是**加强产业链上下游的协同。产业联盟和行业协会应组织芯片、算法、系统、应用服务等各环节企业，共同提炼互操作性需求，形成统一的产业立场，加强互操作性。**三是**引导企业将标准符合性测试前置于产品研发阶段，降低合规成本，推动标准落地。**四是**支持构建标准验证与推广生态。产业界应联合科研机构，共同加强测试方法、验证工具及认证体系建设，形成标准推广与产业发展的良性合力。

## （四）科研机构加深理论基础研究，支撑前沿探索与人才培养

**科研机构与学术界**是标准创新的思想库和人才库，应致力于为标准化工作提供坚实的理论根基与前瞻洞察。**一是**深化标准相关基础科学问题研究。鼓励科研机构对人工智能技术路径、安全可信、垂类应用等关键领域进行长期、深入的理论研究，并将坚实的科学发现转化为标准中可用的模型、方法和度量指标，提升标准的技术深度。**二是**构建"科研-标准"双向转化通道。支持学术界密切关注国际标准讨论中暴露出的技术瓶颈，将其设立为重要科研课题；同时，建立机制将前沿研究成果及时、主动地向标准组织报告，启发新的工作方向。**三是**加强标准化人才培养与学科建设。推动在计算机科学、公共政策、伦理学等相关学科中开设人工智能标准与治理课程，培养学生的标准化素养，为未来储备既懂技术又懂规则、具备国际视野的复合型人才。

**人工智能标准化是推动全球协同治理和产业可持续发展的重要基石。**自2024年乌镇峰会以来，世界互联网大会人工智能专业委员会依托标准推进计划，汇聚国际组织及各国产学研代表，推动在标准现状、发展态势、困难挑战等领域合作对话，共建人工智能标准化国际交流互鉴平台。未来，大会将继续秉持共商共建共享基本原则，与国际社会一道，持续推进负责任的人工智能标准联合研究、互鉴合作、应用推广与迭代更新，共同推动全球人工智能标准发展进程。

# 05 附录：全球负责任的人工智能标准实践案例

## （一）可持续发展推动以人为本和智能向善

**ITU通过AI for Good倡议推动全球负责任人工智能标准化实践。**作为联合国主办的全球性平台，ITU联合40多个联合国机构共同发起"人工智能向善"（AI for Good）全球倡议，旨在识别和推广有助于实现联合国可持续发展目标的人工智能应用，构建全球技能与标准体系，促进多方合作。该倡议通过国际人工智能标准交流平台汇聚全球领先的标准化组织和专家，推动人工智能伦理、透明度、可问责性、隐私保护等领域的标准制定与实践落地。此外，ITU还主办了"人工智能向善全球峰会"，旨在通过全球范围内的项目加速人工智能技术在健康、教育、环境等领域的应用，推动全球可持续发展。通过这些举措，ITU为构建可信、可持续的人工智能生态系统提供了重要支撑，促进了全球范围内的技术创新与社会责任的有机融合。

**IEEE通过CertifAIEd项目推动人工智能伦理培训和标准化实践。**IEEE CertifAIEd是面向全球的实时在线培训与认证项目，旨在帮助学员理解和应用人工智能伦理标准，提升自主智能系统的负责任开发与使用能力。该项目不仅普及人工智能伦理的基本原则，还提供了完整的评估方法论，指导学员如何将其应用于人工智能产品和服务的专业评估之中。在认证框架中，IEEE明确提出了透明度、可问责性、算法偏见防控与隐私保护四大标准要素：强调系统设计中价值观与决策的公开披露，要求组织对人工智能系统结果承担责任，防止因算法缺陷导致的不公正结果，并尊重个人与群体的隐私与尊严。通过体系化的培训和认证，IEEE推动人工智能伦理原则在产业中的落地转化，为构建兼顾技术进步与社会责任的可持续人工智能发展模式提供了重要支撑。

**科大讯飞推动人工智能应用服务标准化赋能教育可持续发展。**科大讯飞深度参与《教学应用质量规范 核心框架》《区域教育数字化转型成熟度》等教育信息化标准研制工作，为人工智能应用教育领域的效果提供了科学的评价方法和基准；同时，积极践行标准化成果，率先在教育行业开展人工智能领域技术试点示范，将《教育人工智能大模型接口规范》等标准与智慧教育产品深

度结合。其自主研发的"高中数学智能教师系统"通过标准化接口实现个性化学习方案输出，服务全国32个省份及海外市场。此外，科大讯飞还将《移动学习终端功能要求》《教师数字素养》《测试试题信息模型》等标准应用于学习机、智能黑板等终端设备，为标准实施提供了示范性应用案例。这些标准化成果不仅规范了人工智能教育应用，还为教育管理部门评估应用效果提供了科学依据。

## （二）能力建设提升产业国际标准制定水平

**全国网络安全标准化技术委员会（TC260）推出《人工智能安全治理框架》，推动中国人工智能安全治理标准化探索。**落实《全球人工智能治理倡议》，国家互联网信息办公室指导TC260发布《人工智能安全治理框架》1.0版和2.0版，提出人工智能安全治理框架性方案，并动态调整风险分类，优化完善防治措施，推动人工智能协同共治、普惠共享。组织研制《人工智能安全标准体系》，采用国家标准、技术文件和实践指南相结合的形式灵活推动标准工作，为国家人工智能安全治理提供技术支撑。围绕我国人工智能安全治理急需，发布《生成式人工智能服务安全基本要求》《生成式人工智能预训练和优化训练数据安全规范》《生成式人工智能数据标注安全规范》《人工智能生成合成内容标识方法》等国家标准，组织研制人工智能应用安全分类分级安全能力成熟度评估、涉及未成年人应用等国家标准。围绕人工智能应用安全重点需求，发布《政务大模型应用安全规范》技术文件，研制金融、交通、广电、教育、医疗等行业领域安全指引，促进人工智能健康发展与安全应用。

**中国信息通信研究院通过标准协同机制助推人工智能产业健康有序发展。**中国信息通信研究院长期深耕人工智能标准化工作，国家标准方面，承担工业和信息化部人工智能标准化技术委员会秘书处职能，不断强化人工智能标准化体系建设，自2025年成立以来，制定发布《工业和信息化部人工智能标准化技术委员会2025年标准制定指南》《工业和信息化领域人工智能安全治理标准体系建设指南(2025)》，推进26项人工智能行业标准立项，联合CCSA推动发布20项人工智能行业标准，涵盖大模型、具身智能等热点方向，面向860余名标准化工作者开展培训。国际标准方面，承担国际电信联盟办公室秘书处职能，同时，作为ITU-T SG21国内对口组组长单位，已累计推动国际标准100项，其中在研项目47项，发布项目53项，实现了从芯片、软件框架、工程平台、大模型及上层应用的全体系布局，也引领性的发布了大模型评测、人工智能平台、生成式人工智能、数字人等四个代表性标准体系。中国信息通信研究院通过系统推进人工智能标准体系建设与治理实践，为构建可信、包容、可持续的人工智能治理生态提供支撑。

**中国电信积极推动信息通信领域人工智能标准化国际合作。**2024年11月的世界电信标准化全会（WTSA-24）期间，中国电信共牵头起草7项亚太地区提案，接近中国提案总数的一半。其中，由中国电信、中国信息通信研究院等产业主体牵头的《关于支持电信/ICT领域人工智能技术标准化活动》新决议获得大会正式通过。此项决议核心意义在于将人工智能标准在电信行业的应用正式纳入全球制度化框架，为未来相关国际规则的制定和技术工作的协同搭建了关键平台，有效促进了全

球产业的开放合作与健康发展。该决议融入“以人为本”与“智能向善”的理念，为构建负责任的人工智能发展生态提供了重要指引，倡导技术进步应真正服务于全球可持续发展事业。此外，中国电信专家还在多项国际提案中发挥主导作用，涉及加强标准化组织间的协作、呼叫者号码标识、打击伪造和失窃电信设备以及元宇宙等新兴议题，推动了更广泛领域的技术共识构建。

**商汤科技深度参与各国际标准组织人工智能标准制定。**作为中国人工智能产业的重要代表，商汤科技长期致力于国际标准的研发与推广。在ISO中，商汤多位专家深度参与WG1基础组和WG3可信赖组，贡献于人工智能基础技术、安全可信及管理认证相关标准的制定；在IEC中，参与生物数字融合系统评估组（SMB SEG12）工作，推动跨学科融合标准探索；在ITU中，积极参与SG16相关的基础技术及数字人标准研制，支撑数字人应用的国际协作框架；在IEEE中，其牵头或主导多个标准工作组，包括P3110计算机视觉、P2945人脸识别、P3375人工智能芯片、P2048.101增强现实、P2048.111空间计算、P2048.121数字人等项目，推动计算机视觉、智能硬件与沉浸式交互等前沿领域的技术共识形成。商汤科技已累计深度参与20余项国际标准项目，涵盖大模型、智能计算、人工智能管理体系、数据质量、数字人等关键领域，也是首个提出人脸识别国际标准提案的中国企业。

## （三）普惠创新助力技术开放共享广泛应用

**京东JoyAI大模型加速产业走向深度应用。**京东积极参与人工智能国家标准体系建设，参与了《人工智能 可信赖 第1部分：通则》《人工智能 隐私计算通用框架》等多项国家标准的制定。在京东JoyAI大模型的设计与研发过程中，注重贯彻相关标准要求，通过标准牵引实现了技术普惠与场景包容。京东JoyAI大模型涵盖语言、语音、图像、视频、数字人等多种模态，通过动态分层蒸馏、跨领域数据治理等创新技术，兼顾了“大而精”。在客服应答等需快速响应的场景中，系统自动启用轻量化推理路径，实现毫秒级反馈；在供应链优化、医疗诊断等复杂决策场景中，则启动深度思考模式，依托1280K的超长上下文窗口，完成多维度信息关联与逻辑推演，在长文本1280K大海捞针评测中准确率接近100%。基于人工智能标准引领，以及京东自有的供应链场景优势，京东JoyAI大模型已深入零售、物流、健康、金融、工业等诸多领域，服务超百万商家，深度应用于数百个细分业务场景，形成标准驱动技术迭代以及场景验证普惠价值的闭环生态。

**京东数字人规模化革新电商直播。**京东数字人遵循国际标准，首批通过中国信通院“数字人系统基础能力评测”。依托“建模-驱动-渲染”一体化的端到端技术路线，京东攻克数字人生成幻觉难题，在形象逼真度、语音自然度、交互智能性等多维度技术达到国际领先水平，荣获中国智能科学技术最高奖“吴文俊人工智能科学技术奖”特等奖。京东将单个数字人生产成本从数万元降低至两位数，较真人拍摄模式成本降幅超90%，标志着数字人从专业工具正式成为普惠生产力。京东数字人已支持超2万家品牌在京东及其他平台上开展直播，覆盖美妆、珠宝首饰、3C数码、家电家居、医疗健康、食品酒水、汽车等全品类，累计带动GMV超百亿元，验证了标准对“低成本、高并发、广覆盖”商业普惠的放大效应，为行业提供了可大规模复制的落地范式。

**vivo深度布局人工智能手机、端侧大模型、智能体标准，提升手机个人化智能体验。**vivo倡导"个人化智能"的理念，通过1+X+N的战略布局为用户提供个人化、温暖的智能体验，打造专属助理，结合底层系统能力打造全局统一人工智能工具及三方生态服务的构建。以此为基础，深度参与CCSA、MIIT/TC1、电信终端产业协会（TAF）、中国人工智能产业发展联盟（AIIA）等标准与产业组织关于人工智能手机、端侧大模型和智能体的标准化工作，重点牵头和参与《智能体技术要求与评估方法 第7部分：手机智能体》《智能终端意图框架总体技术要求》《智能终端意图框架接口技术要求》等标准制定。在业务实践中，注重贯彻相关标准要求，通过标准牵引实现了技术普惠，同时收获了多项行业肯定，包括：业内首家通过中国信通院手机智能体评估，并获得当前最高评级（4+级）；蓝心语音大模型顺利通过可信AI语音大模型专项评估（4+级）；vivo蓝心端侧大模型行业首家获得端侧大语言模型能力证书。

**长安汽车以国际标准为牵引，构建可复现、可评测、可落地的人机交互与感知标准体系。**在《车载多模态语音交互系统功能架构》中，明确语音、唇动与手势的混合决策链路和模块接口，提出误唤醒及模糊指令处理的标准化建议，提升模块互操作性与工程可移植性。在《车载智能体增强语音对话系统框架与要求》中，标准化对话代理的基础框架和核心能力，覆盖上下文理解、自治规划、工具调用及个性化记忆，并强调可控性与安全约束，实现类人流畅可审计交互。在《车载视频处理组件评估框架》中，提出主客观质量、时延稳定性、场景鲁棒性和能效等指标与测试流程，覆盖驾驶及舱内交互关键场景。三项标准形成"架构—能力—评价"闭环，通过统一术语、数据与接口，支撑复杂场景下负责任人工智能的量化、可追溯和可持续应用，为监管合规和产业推广提供实践依据。

**安谋科技自研神经网络处理器技术方案（NPU IP）标准化应用推动端侧人工智能落地。**安谋科技积极参与全国汽车标准化技术委员会（TC114）、CCSA、MIIT/TC1端侧人工智能标准，参与编写《汽车智能座舱计算芯片技术要求及试验方法》《人工智能芯片基准评估方法》等行业标准。新一代"周易"NPU IP采用面向大模型特性优化的架构设计，针对包括智能座舱、驾驶辅助等在内的多种应用场景进行了深度性能提升，以更好地满足汽车智能化对高算力与高能效的严格要求。"周易"NPU IP提供完善的Compass软件开发套件及成熟的网络部署工具AIPULLM，能够支持多种主流模型的部署需求。"周易"NPU IP已成功集成于芯擎科技"龙鹰一号"和芯驰科技X9智舱处理器等多款芯片中，广泛应用于车载智能场景。其中，芯擎科技"龙鹰一号"已在多家车厂定点达数十款主力车型，累计出货超百万颗。

**上海诺基亚贝尔生成式人工智能平台提升标准化工作效率。**作为旗下首款私域部署平台，FusionDeep凭借本地算力和大语言模型确保数据安全合规，并支持敏捷迭代与多智能体融合，能够快速响应定制化需求，全面赋能企业业务转型。FusionDeep核心应用包括私域知识库聊天机器人、文档助手与聊天式报表，支持一键总结长篇报告、智能翻译文档，并提取会议纪要。在标准化方面，FusionDeep开发了多项提升标准化工作质量与效率的功能，如标准文稿的修订和润色，以

及标准报告的生成。此外，FusionDeep还提供API接口，可与企业现有业务系统无缝对接，拓展更多定制化场景。FusionDeep正在助力企业加速人工智能驱动的数字化转型，释放新的业务增长动力，开创智能化的未来。

## （四）风险管理划定人工智能发展责任边界

**ISO和IEC推出全球首个人工智能管理体系标准倡导风险导向的负责任开发和使用。**ISO/IEC 42001:2023《人工智能管理体系》采用"计划—执行—检查—行动"（PDCA）方法，帮助组织在人工智能系统的整个生命周期内识别、评估和应对风险。关键要求包括：一是系统识别和评估人工智能相关的风险，包括伦理、透明度和持续学习等方面，制定相应的应对措施；二是涵盖人工智能系统从设计、开发、部署到退役的全过程，确保每个阶段的风险得到有效管理；三是对外部供应商提供的人工智能产品或服务进行监督，确保其符合组织的风险管理要求。通过实施该标准，组织能够建立透明、可追溯和可靠的人工智能管理体系，提升公众和利益相关者的信任，促进人工智能技术的健康发展。

**之江实验室专业领域大模型通过标准应用保障负责任的人工智能实践。**GeoGPT是面向全球开放的非营利地学大模型平台，自诞生以来高度关注数据开放、模型透明、伦理治理、偏见防控、隐私保护等问题。GeoGPT项目成立了国际化治理委员会，其由来自不同国家的地球科学、人工智能、数据治理、法律与伦理等领域专家组成，秉持信任、透明、遵循伦理和独立性的原则对平台进行监督，并提供战略指导、伦理监督、风险管控与开放性保障，确保大模型技术创新和社会责任之间的平衡。GeoGPT项目结合ITU-T Y.3172《未来网络中机器学习的架构框架》、ISO/IEC 42001:2023《人工智能管理体系》、ISO/IEC 23894:2023《人工智能风险管理》等国际标准，结合联合国教科文组织《人工智能伦理建议书》、欧盟《人工智能法》、美国NIST《人工智能风险管理框架》等国际倡议、区域法规与合规指南，构建并实施针对平台的人工智能风险评估框架。在此过程中，GeoGPT不仅完成了标准对接与验证，还探索了标准在人工智能治理与风险管控应用中的落地路径，实现负责任原则在大模型研发与部署中的有效转化，并通过实践反馈促进标准持续完善。

## （五）安全可靠应对技术内生与社会衍生挑战

**中国电信深度鉴伪专家系统为应对内容安全风险提供可验证的技术路径。**该系统依托大模型框架，融合多模态信息，能够在视频通话、身份认证等高敏感场景中实现实时伪造检测，检测准确率和稳定性均处于行业领先水平。目前，该系统已在国际国内多项权威测试中验证其可行性，并通过行业机构认证。其鉴伪能力已推广至全国多个省份和部分海外市场，支撑实名认证、移动通信和跨境业务的安全应用，有效提升了社会公众对人工智能系统的信赖度。基于此，中国电信在ITU-T SG21中主导制定《视频真实性检测服务的评估标准》，明确视频真实性检测服务的技术要求、评估类别、关键性能指标及测试方法，为金融、安防等关键领域的应用方提供客观、可复现的评价基

准。此外，在ISO/IEC JTC1 SC29中，中国电信担任多媒体内容真实性标准专题组联合主席，联合Dobby和Fraunhofer HHI编制《信息技术 媒体真实性 词汇》标准，定义媒体真实性领域的相关术语，为全球技术研发、产业协作和政策法规制定奠定了沟通基石。

**360智能体安全卫士全面提升智能体领域安全可靠性。**360充分发挥跨领域的技术优势，系统参与编写《人工智能 大模型 第1部分：通用要求》《人工智能 大模型 第2部分：评测指标与方法》《人工智能 大模型 第3部分：服务能力成熟度评估》等人工智能大模型国家标准，聚焦技术规范与评估体系顶层设计。同时，360面向智能体的全生命周期构建安全防护体系，为行业智能化升级提供坚实安全基座，助力打造人工智能发展高地。一是依托于云端安全运营机制，逐步打造智能体供应链威胁监测体系。基于成熟的云端安全运营体系与沙箱检测能力，打造智能体供应链云端威胁监测能力，维护智能体工具和服务的白名单。二是基于成熟的端侧安全防护框架，构建智能体本地化防护套件。实现端侧的工具鉴定、服务风险监测等，与云端运营体系配合实现体系化防护。三是打造智能体行为沙箱，有效保障运行时安全。通过构建端侧的隔离运行环境，实现对智能体执行过程中全生命周期行为的安全评估，及时发现恶意行为，实现动态安全防护。

**奇安信结合先进网络安全技术助力人工智能安全标准制定。**近年来，奇安信围绕SecAI安全检测方法、安全大模型技术以及基于流量检测识别和监测审计的大模型安全防护技术，参与《网络安全技术 人工智能计算平台安全框架》《信息安全技术 互联网信息服务安全通用要求》等已发布国家标准制定，参与《网络安全技术 人工智能安全能力成熟度评估方法》《网络安全技术 人工智能应用安全分类分级方法》《信息安全技术 互联网信息服务深度合成安全规范》以及《网络安全标准实践指南——生成式人工智能服务安全应急响应指南》等国家标准和实践指南研制，并参与中国人工智能安全能力评估方法、分类分级方法以及互联网信息深度合成安全技术要求等政策研究，为推动形成人工智能安全标准体系提供助力。此外，奇安信还积极参与《电信网与互联网安全大模型 数据安全领域》等行业标准制定，推动安全大模型技术在具体行业领域应用与推广。

## （六）数据治理保障数据质量流通与价值实现

**科大讯飞主导制定首个中国牵头的人工智能数据国际标准。**在技术创新和应用落地的过程中，科大讯飞一直高度重视数据治理，严格控制数据质量，以高质量数据驱动人工智能技术在更广泛领域的应用，促进人工智能产品和服务可用、可靠、可信。在标准化实践中，科大讯飞主导制定首个由中国牵头的人工智能领域数据国际标准ISO/IEC 5259-4:2024《人工智能 分析和机器学习的数据质量 第4部分：数据质量过程框架》，标志着中国在全球人工智能领域数据标准化进程中迈出了坚实的一步。该标准适用于监督学习、无监督学习、半监督学习、强化学习以及分析中所用数据的质量控制，为各类组织提供了通用方法和指南。这些指导原则覆盖了数据的全流程，如数据的获取、准备、标注、评估、使用等。其落地将有助于推动行业建立更加透明和规范的数据治理机制，支撑人工智能产业的高质量和可持续发展。

## （七）隐私保护强化个人权益保护与社会信任

**荣耀通过人脸合成反诈检测系统探索隐私保护标准化路径。**面对人工智能换脸诈骗等新型安全威胁，荣耀依托终端侧人工智能技术，创新推出人脸合成反诈检测系统，实现对用户视频画面内容的本地化自主识别，能够精准检测生成合成人脸，并在发现可疑迹象时即时向用户发出风险预警，有效帮助消费者防范新型诈骗风险，切实保障个人隐私与财产安全。该系统已形成完整的全链路反诈能力，兼顾隐私保护与安全应用，获得产业界高度认可。2024年，荣耀"智能终端AIGC人脸合成反诈检测系统"荣获中国网络安全创新大赛"最具投资价值奖"，并入选中国网络安全协会评选的"人工智能创新应用典型案例"，在主题展上作为综合性成果进行展示。在实践基础上，荣耀还主导发起《移动智能终端人脸合成检测技术能力评估方法》标准立项，系统提出人脸合成检测的技术能力框架、评估指标与评估方法，为行业研发与评估终端侧人脸合成检测技术提供了标准化参考。


世界互联网大会
World Internet
Conference