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）

（征求意见稿）

为深入贯彻落实党中央、国务院关于加快发展人工智能的部署要求，贯彻落实《国家标准化发展纲要》《全球人工智能治理倡议》《国家人工智能产业综合标准化体系建设指南（2024）》，根据工业和信息化部人工智能标准化技术委员会审议通过的《人工智能标准化技术委员会标准体系（2025年）》，进一步加强人工智能安全领域标准化工作系统谋划，加快构建保障人工智能产业高质量发展和实现高水平安全的标准体系，夯实标准对推动技术进步、促进企业发展、引领产业升级、保障产业安全的支撑作用，更好推进人工智能赋能新型工业化，加速迈向制造强国和网络强国，特制定本指南。

一、背景情况

习近平总书记高度重视统筹人工智能发展和安全，强调“要加强人工智能发展的潜在风险研判和防范，维护人民利益和国家安全，确保人工智能安全、可靠、可控”。党的二十届三中全会明确提出要“建立人工智能安全监管制度”。习近平总书记的系列重要指示批示精神为开展人工智能发展和安全工作指明了方向，提供了根本遵循。

近年来，我国人工智能产业在技术创新、产品创造和行业应用等方面实现快速发展，在人工智能安全治理标准制定

方面也取得了一定进展，但仍面临数据安全、网络安全等方面挑战。立足长远，构建人工智能安全治理标准体系迫在眉睫。**一是**提升人工智能企业安全能力建设，培育壮大人工智能产业，服务本土企业长远发展。**二是**培育人工智能安全技术及产品应用，提升人工智能赋能安全能力，为人工智能技术高质量应用、数字经济和实体经济融合提供坚实支撑。**三是**明确人工智能技术研发、产品迭代安全基线，促进企业在安全框架内创新发展，增强产业主体的使用信心，营造安全的发展环境，同时促进技术共享、合作共赢，加速技术成果在安全可靠的前提下落地转化。**四是**提升我国人工智能技术产品的国际竞争力，通过将我国先进的人工智能安全防护理念、实践成果推向国际社会，吸引国际合作资源，为本土企业拓展海外业务筑牢根基。

二、总体要求

（一）指导思想

本指南以习近平新时代中国特色社会主义思想为指导，全面贯彻党的二十大和二十届二中、三中全会精神，以人工智能赋能新型工业化为主线，遵循统筹发展和安全的基本原则，构建工业和信息化领域人工智能安全治理体系，为产业提供人工智能安全的技术和行动指引。以高水平安全促进高质量发展，充分发挥产业主体在人工智能安全治理的重要作用，加速培育壮大人工智能产业，构建安全、可靠、繁荣的人工智能产业发展环境。

（二）建设目标

短期目标（1-2 年）：急用先行，快速突破。初步构建工

业和信息化领域人工智能安全标准体系框架，明确标准体系的总体架构、分类和关键标准领域。制定一批急需急用的标准，为行业提供基本的安全规范和技术指导。推动相关标准的试点应用，在部分重点企业和项目中开展标准实施的验证和反馈工作，及时发现和解决标准在应用过程中存在的问题，为标准的完善和推广奠定基础。

长期目标（3-5 年）：全面保障，产业落地。完善工业和信息化领域人工智能安全标准体系，补充和细化各领域的安全标准，涵盖人工智能系统的设计、开发、部署、运行、维护等全生命周期，覆盖工业和信息化领域人工智能应用的各个方面，并促进与国际标准的对接协调。同时，加强标准的推广应用，规范工业和信息化产业主体人工智能应用，提高行业整体的安全水平，为传统产业的高端化升级和前沿技术的产业化落地提供有力保障。

三、建设思路

（一）人工智能安全治理标准体系结构

人工智能安全治理标准体系结构包括“A 治理能力”、“B 基础设施安全”、“C 网络安全”、“D 数据安全”、“E 算法模型安全”“F 应用安全”“G 赋能安全”等 7 个部分，如图 1 所示。

图 1：人工智能安全治理标准体系结构

其中，治理能力标准主要规范人工智能支撑能力和管理能力，为实现人工智能安全治理夯实基础底座。基础设施安全标准主要规范硬件平台、软件平台和智算中心等方面安全，确保人工智能基础设施安全稳定。网络安全标准主要规范网络安全的防护、监测、管理，以及供应链安全、网络安全风险评估，确保网络安全可靠运营。数据安全标准主要规范基础数据服务、训练数据、业务数据等方面安全，明确人工智能数据安全要求。算法模型安全标准主要规范算法、模型等方面安全，保障人工智能技术创新安全可控。应用安全标准主要规范智能产品应用、智能服务应用、智能体、信息等方面安全，为推动产业智能化发展提供安全保障。赋能安全标准主要规范人工智能赋能新型工业化应用、行业应用，以及赋能网络、数据、信息、业务等方面安全，为人工智能赋能

安全提供技术保障。

（二）人工智能安全治理标准体系框架

人工智能安全领域标准体系框架主要由治理能力、基础设施安全、网络安全、数据安全、算法模型安全、应用安全、赋能安全等 7 个部分组成，如图 2 所示。

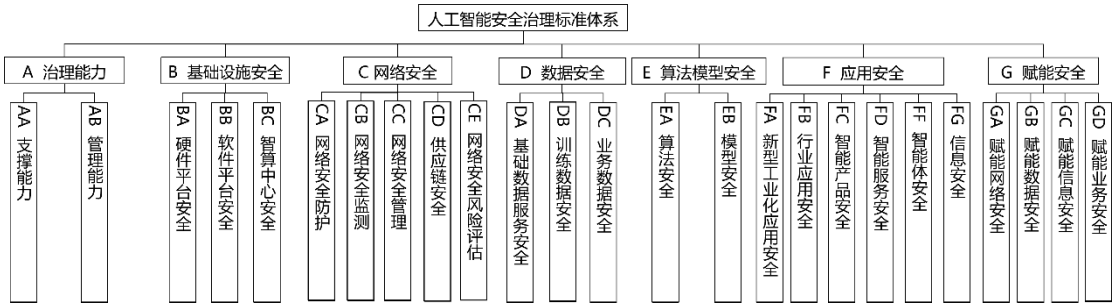

图 2：人工智能安全治理标准体系框架图

四、建设内容

（一）人工智能治理能力

人工智能治理能力标准主要包括支撑能力、管理能力等标准，为安全治理标准体系奠定基础底座。

1. 支撑能力。规范人工智能系统全生命周期的安全治理技术能力要求，包括验证、监测、防护、追溯、证真、鉴伪等安全支撑技术；透明性、可解释性、鲁棒性、公平性、可靠性、可追溯性、隐私保护以及用户权益保障等治理支撑技术。

2. 管理能力。规范人工智能的技术研发和运营服务安全治理要求，包括可信赖研发管理、风险分类分级、风险管理能力、风险影响评估、风险应急响应等标准。

（二）人工智能基础设施安全

人工智能基础设施安全标准主要包括硬件平台安全、软件平台安全、智算中心安全等标准，为人工智能提供基础运营环境安全保障。

1. 硬件平台安全。规范硬件平台的安全设计、生产和测试要求，包括芯片、传感器、计算设备等标准，芯片抗攻击能力、传感器数据完整性、计算设备的物理安全性等能力评估，硬件设备的安全设计原则、安全功能模块及接口规范等安全要求，硬件安全性能的物理攻击模拟、故障注入测试等测评方法。

2. 软件平台安全。规范软件平台的安全性技术要求，包括系统软件的漏洞风险、开发框架的安全性、软件供应链的完整性能力评估，软件开发生命周期中的安全规范、安全编

码标准及漏洞管理流程的安全框架要求，软件平台的访问控制、权限管理、数据保护及安全审计要求，代码审计、渗透测试、风险测试方法和评估标准。

3. 智算中心安全。规范面向人工智能的算力中心安全要求，包括面向人工智能的大规模计算集群、新型数据中心、智算中心、算力网络等基础设施的安全要求，访问控制、加密传输等网络安全、数据安全、信息安全技术要求标准，以及安全防护、应急响应等管理标准。

（三）人工智能网络安全

规范人工智能网络安全要求，包括网络安全防护、网络安全监测、网络安全管理、供应链安全等。

1. 网络安全防护。规范人工智能相关产品、服务的网络安全防护要求，结合全生命周期不同阶段存在的网络安全风险，明确需要的安全措施，例如攻击防范、漏洞管理、模型更新安全、安全应急响应等。

2. 网络安全监测。规范开展人工智能网络安全监测/监管的安全要求，包括通用要求、技术要求、测试方法、接口规范等。

3. 网络安全管理。规范人工智能产品或服务运营单位开展网络安全定级备案、资产安全管理、漏洞安全管理等安全管理工作时需遵循的要求，规范网络安全风险评估相关要求。

4. 供应链安全。规范人工智能软硬件供应链在安全方面的要求，包括供应商评估、软硬件供应管理、风险识别和防范等方面。

5. 网络安全风险评估。规范人工智能场景开展网络安全

风险评估应遵循的安全要求，包括网络安全评估指南、实施方法、评估要求等。

（四）人工智能数据安全

规范人工智能数据安全要求，包括训练数据安全、业务数据保护、基础数据服务安全等。

1. 基础数据服务安全。规范人工智能技术相关基础数据服务安全要求，包括数据库存储、数据清洗、数据标注、数据分析、数据共享等服务安全。

2. 训练数据安全。规范人工智能预训练和优化训练数据及其处理活动的安全要求，包括数据通用安全以及数据收集、传输、存储、使用等环节安全，训练数据安全防护、安全标注、质量治理、重要数据识别、安全互操作等标准。

3. 业务数据保护。规范运用人工智能技术相关业务的数据安全保护要求，包括数据完整性、保密性、可用性等。规范人工智能技术、产品或服务应用过程中涉及业务数据的安全要求，包括重要数据识别、数据分类分级、全生命周期安全等标准。

（五）人工智能算法模型安全

人工智能算法模型安全标准主要包括算法安全、模型安全等标准。

1. 算法安全。规范深度学习算法、机器学习算法等人工智能算法安全具体要求，研制算法设计的安全可信、算子库安全等标准，算法可靠性、可控性、准确性、可解释性等具体要求。

2. 模型安全。规范模型训练、推理、部署等环节的安全

要求，包括大模型安全要求、混合模型安全要求、模型部署安全、模型安全评测指标与方法、多模型协同调度安全等标准，人机混合增强智能、群体智能、跨媒体智能、具身智能等技术应用中的模型安全要求。

（六）人工智能应用安全

人工智能应用安全结合新型工业化、智能网联汽车、生成式人工智能、生物特征识别、智能体等典型应用，提出新型工业化应用安全、行业应用安全、智能产品应用安全、智能服务应用安全，以及智能体安全等新产品形态安全的要求和评估方法。

1. 新型工业化应用安全。规范人工智能技术在制造业全流程智能化的应用安全要求，包括工业研发设计、中试验证、生产制造、营销服务、运营管理等制造业全流程网络安全防护、评测指标与方法、安全运营、风险评估等标准。

2. 行业应用安全。规范人工智能在不同行业中的应用安全要求，涵盖人工智能行业应用安全开发、监控审计等能力评估，人工智能应用的安全架构框架要求，应用数据隐私保护、合规性要求规范，人工智能应用的安全测试评估方法。

3. 智能产品安全。规范人工智能技术和产品在具体应用场景下的安全要求，包括智能机器人、智能运载工具、智能移动终端、数字人、智能系统、生物特征识别终端等智能产品安全防护与检测标准。

4. 智能服务安全。规范面向特定业务场景提供的人工智能服务安全保障要求，包括智能软件开发安全、智能设计安全、模型即服务安全、生成式人工智能服务安全、生物特征

识别服务安全、智能产品辅助操作安全、智能服务安全运营、智能服务安全风险评估、智能服务安全保障能力评价等标准。

5. 智能体安全。规范以通用大模型为核心的智能体实例安全要求，包括智能体内生安全、智能体数据接口安全、人机协作安全、智能体自主操作安全、多智能体协作安全等标准。

6. 信息安全。规范深度合成、生成式人工智能等在应用过程中的内容安全等方面的要求，制定安全规范和评估方法。规范利用深度合成和生成式人工智能技术生成内容、算法审核及评估、合成鉴伪检测技术等内容管理的安全要求，包括生成内容的模型引导、模板制作、内容审核等方面的技术规范要求，算法审核及评估的算法机制机理审核、信息发布审核、涉生成违法违规内容审查等能力评估，合成鉴伪检测技术的测评要求、评测方法和工具箱、指标体系等评价标准。

（七）人工智能赋能安全

人工智能赋能安全标准主要包括人工智能赋能网络、数据、信息、业务和其他安全等方面的要求。

1. 赋能网络安全。指导应用人工智能提升网络安全的效能。包括安全大模型底座，网络安全大模型能力要求，人工智能赋能网络安全产品、网络安全运营、智能问答、加密流量监测、钓鱼邮件分析、暴露面监测、威胁情报智能挖掘、攻防演练靶场、恶意程序分析识别、WEB 攻击检测、渗透测试、恶意代码检测、异常行为分析、智能安全运营等网络安全领域的技术指南、能力要求、评估方法等。

2. 赋能数据安全。指导应用人工智能提升数据安全的效

能。包括人工智能技术应用于数据收集、数据分类分级、数据标注、数据存储等关键环节中，数据资产识别、数据资产血缘分析、数据合规检测、数据风险评估、数据暴露面检测、数据访问异常识别、数据泄露溯源、数据安全防护、API 安全防护、数据安全审计和其他数据安全领域的技术指南、能力要求、评估方法等。

3. 赋能信息安全。指导应用人工智能技术，提升信息安全的效能。包括人工智能技术应用于深度伪造内容检测、互联网舆情监测、不良信息识别、反诈和其他信息安全场景的技术指南、能力要求、评估方法等。

4. 赋能业务安全。指导应用人工智能技术，提升业务安全的效能。包括人工智能技术应用于业务安全智能风控、支付安全、业务安全合规、业务事后审理能力、黑灰产监控能力和其他业务安全场景的技术指南、能力要求或评估方法等。

附件 1：

现有人工智能安全治理领域标准清单

截至 2025 年 2 月，在人工智能安全治理领域，全国信息技术标准化技术委员会（TC28）、全国网络安全标准化技术委员会（TC260）、中国通信标准化协会（CCSA）等技术委员会或工作组，共计发布 55 项国家标准与行业标准。当前，我国在人工智能安全治理标准制定方面具备一定基础，但标准主要呈现点状化、分散性、交叉性等特点，部分人工智能关键技术环节和新兴应用场景标准缺失，尚无法满足快速发展的技术需求。

序号	类型	归口组织	标准编号	标准名称	状态	分类
人工智能治理能力						
1	国家标准	全国网络安全标准化技术委员会	GB 45438-2025	网络安全技术 人工智能生成合成内容标识方法	即将实施	安全
2	国标计划	全国网络安全标准化技术委员会	20240898-T-469	网络安全技术 数字水印技术实现指南	正在审查	安全
3	行标计划	中国通信标准化协会	2024-1352T-YD	人工智能安全治理 生成式人工智能图像检测服务系统能力要求	在研	信息传输、软件和信息技术服务业
4	国家标准	全国信息技术标准化技术委员会	GB/T 45081-2024	人工智能 管理体系	现行	管理
5	国标计划	全国信息技术标准化技术委员会	20240562-T-469	人工智能 可信赖 第 1 部分: 通则	正在起草	基础
6	国标计划	全国信息技术标准化技术委员会	20231740-T-469	人工智能 风险管理能力评估	正在审查	基础
7	行标计划	中国通信标准	2024-1353T-YD	人工智能安全	在研	信息传输、

		化协会		治理 系统风险管理能力要求		软件和信息技术服务业
8	行标计划	中国通信标准化协会	2024-0592T-YD	智能风控平台指标要求和评估方法	在研	信息传输、软件和信息技术服务业
9	行标计划	中国通信标准化协会	2023-0041T-YD	人工智能开发平台通用能力要求 第2部分：安全要求	在研	信息传输、软件和信息技术服务业
人工智能基础设施安全						
10	国标计划	全国网络安全标准化技术委员会	20230249-T-469	网络安全技术 人工智能计算平台安全框架	暂缓	安全
11	行业标准	中国通信标准化协会	YD/T 4255-2023	算力网络 总体技术要求	现行	信息传输、软件和信息技术服务业
12	行业标准	中国通信标准化协会	YD/T 6046-2024	算力网络 算网编排管理技术要求	现行	信息传输、软件和信息技术服务业
13	行业标准	中国通信标准化协会	YD/T 4598.1-2024	面向云计算的零信任体系 第1部分：总体架构	现行	信息传输、软件和信息技术服务业
14	行业标准	中国通信标准化协会	YD/T 4598.2-2023	面向云计算的零信任体系 第2部分：关键能力要求	现行	信息传输、软件和信息技术服务业
15	行业标准	中国通信标准化协会	YD/T 4598.3-2023	面向云计算的零信任体系 第3部分：安全访问服务边缘能力要求	现行	信息传输、软件和信息技术服务业
16	行业标准	中国通信标准化协会	YD/T 4598.5-2024	面向云计算的零信任体系 第5部分：业务安全能力要求	现行	信息传输、软件和信息技术服务业
17	行业标准	中国通信标准化协会	YD/T 4598.6-2024	面向云计算的零信任体系 第6部分：数字身	现行	信息传输、软件和信息技术服

				份安全能力要求		务业
18	行标计划	中国通信标准化协会	H-202304046405	基于零信任的算力网络安全技术要求	在研	信息传输、软件和信息技术服务业
人工智能数据安全						
19	国家标准	全国信息技术标准化技术委员会	GB/T 42755-2023	人工智能 面向机器学习的数据标注规程	现行	基础
20	国标计划	全国信息技术标准化技术委员会	20242095-T-469	网络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	正在审查	安全
21	国标计划	全国网络安全标准化技术委员会	20242097-T-469	网络安全技术生成式人工智能数据标注安全规范	正在审查	安全
22	行标计划	中国通信标准化协会	H-202410309647	电信网和互联网人工智能数据安全通用要求	在研	信息传输、软件和信息技术服务业
23	行标计划	中国通信标准化协会	H-202310167087	电信和互联网大规模预训练模型安全指南	在研	信息传输、软件和信息技术服务业
人工智能算法模型安全						
24	国家标准	全国网络安全标准化技术委员会	GB/T 42888-2023	信息安全技术机器学习算法安全评估规范	现行	安全
25	国家标准	全国信息技术标准化技术委员会	GB/T 45225-2025	人工智能 深度学习算法评估	现行	基础
26	行标计划	中国通信标准化协会	H-202307056802	电信和互联网人工智能算法安全要求	在研	信息传输、软件和信息技术服务业
27	行标计划	中国通信标准化协会	H-202302286191	信息通信行业人工智能算法安全评估指南	在研	信息传输、软件和信息技术服务业
人工智能应用安全						

28	国家标准	全国道路交通管理标准化技术委员会	GB/T 43766-2024	智能网联汽车运行安全测试技术要求	现行	方法
29	国家标准	全国道路交通管理标准化技术委员会	GB/T 44850-2024	智能网联汽车运行安全测试项目和方法	即将实施	方法
30	国家标准	全国网络安全标准化技术委员会	GB/T 41819-2022	信息安全技术人脸识别数据安全要求	现行	安全
31	国家标准	全国网络安全标准化技术委员会	GB/T 40660-2021	信息安全技术生物特征识别信息保护基本要求	现行	安全
32	国家标准	全国网络安全标准化技术委员会	GB/T 38542-2020	信息安全技术基于生物特征识别的移动智能终端身份鉴别技术框架	现行	安全
33	国标计划	全国网络安全标准化技术委员会	20241752-T-469	网络安全技术生成式人工智能服务安全基本要求	正在审查	产品
34	国标计划	全国智能技术社会应用与评估基础标准化工作组	20242884-T-469	生成式人工智能技术应用社会影响 服务提供者合规管理指南	正在起草	管理
35	国标计划	全国智能技术社会应用与评估基础标准化工作组	20242891-T-469	生成式人工智能技术应用社会影响 评估指南	正在起草	管理
36	国标计划	全国汽车标准化技术委员会智能网联汽车分会	20243203-T-339	智能网联汽车数据安全管理体系规范	正在起草	管理
37	行业标准	中国通信标准化协会	YD/T 4960-2024	移动智能终端可信人工智能安全指南	现行	信息传输、软件和信息技术服务业
38	行业标准	中国通信标准化协会	YD/T 4679-2024	基于人工智能的诈骗电话号码识别技术要求	现行	信息传输、软件和信息技术服务业

39	行业标准	中国通信标准化协会	YD/T 4994-2024	移动智能终端人工智能应用的个人信息保护技术要求及评估方法	现行	信息传输、软件和信息技术服务业
40	行标计划	中国通信标准化协会	2023-0039T-YD	面向人脸识别系统的人脸信息保护基础能力要求	在研	信息传输、软件和信息技术服务业
41	行标计划	中国通信标准化协会	H-202308286945	生成式人工智能服务的个人信息保护规范	在研	信息传输、软件和信息技术服务业
42	行标计划	中国通信标准化协会	H-202403188928	生成式人工智能终端 安全技术要求	在研	信息传输、软件和信息技术服务业
43	行标计划	中国通信标准化协会	H-202406179281	生成式人工智能语音助手服务的用户权益保护要求	在研	信息传输、软件和信息技术服务业
44	国标计划	全国网络安全标准化技术委员会	20240395-T-469	网络安全技术 互联网信息服务深度合成安全规范	正在起草	安全
45	行标计划	中国通信标准化协会	2022-1488T-YD	互联网信息服务提供商音视频深度合成鉴伪检测技术能力评测方法	在研	信息传输、软件和信息技术服务业
46	行标计划	中国通信标准化协会	2022-1490T-YD	基础电信企业深度合成鉴伪检测系统技术要求及接口规范	在研	信息传输、软件和信息技术服务业
47	行标计划	中国通信标准化协会	2023-1328T-YD	图像视频深度合成鉴伪检测算法引擎安全性评测方法	在研	信息传输、软件和信息技术服务业
48	行标计划	中国通信标准化协会	2023-1332T-YD	音频深度合成鉴伪检测算法引擎安全性评测方法	在研	信息传输、软件和信息技术服务业

49	行标计划	中国通信标准化协会	2023-1337T-YD	互联网数据中心免深度合成鉴伪检测可信机房或可信 IP 认定指南	在研	信息传输、软件和信息技术服务业
50	行标计划	中国通信标准化协会	H-202307196818	互联网深度合成信息服务标识通用安全要求	在研	信息传输、软件和信息技术服务业
51	行标计划	中国通信标准化协会	H-202307196819	互联网音视频深度合成工具安全指南	在研	信息传输、软件和信息技术服务业
52	行标计划	中国通信标准化协会	H-202304106474	互联网音视频深度合成工具安全风险评估方法	在研	信息传输、软件和信息技术服务业
人工智能赋能安全						
53	行标计划	中国通信标准化协会	H-202405139134	电信网和互联网安全大模型能力要求 总体框架	在研	信息传输、软件和信息技术服务业
54	行标计划	中国通信标准化协会	H-202405309219	电信网和互联网安全大模型能力要求 网络安全领域	在研	信息传输、软件和信息技术服务业
55	行标计划	中国通信标准化协会	H-202409309580	安全大模型能力要求 数据安全领域	在研	信息传输、软件和信息技术服务业

附件 2：

下一步计划标准清单

序号	任务分类	标准名称	标准类型	归口组织	紧急程度
1	AA	人工智能代码安全检测技术基础能力要求	行业标准	中国通信标准化协会	2 年
2	AA	人工智能生成内容确权与可追溯性基本要求	行业标准	中国通信标准化协会	2 年
3	AA	人工智能 安全治理 大模型安全基准测试总体技术要求	行业标准	中国通信标准化协会	1 年
4	AA	人工智能 安全治理 人工智能生成内容证真技术要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
5	AA	人工智能 安全治理 大模型可解释性技术要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
6	AA	人工智能 安全治理 大模型隐私保护技术要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
7	AB	人工智能 安全治理 可信研发管理基本要求	行业标准	工业和信息化部人工智能标准化技术委员会	1 年
8	AB	人工智能 安全治理 风险分级分类要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
9	AB	人工智能 安全治理 风险影响评估指南	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
10	AB	人工智能 安全治理 人工智能服务透明性及用户权益保障能力要求	行业标准	工业和信息化部人工智能标准化技术委员会	1 年
11	BC	智算中心安全评估规范	行业标准	中国通信标准化协会	1 年
12	BC	面向人工智能的算力中心安全技术要求与测试方法	行业标准	中国通信标准化协会	2 年
13	BC	算力网络安全技术要求与测试方法	行业标准	中国通信标准化协会	3 年
14	CD	人工智能平台供应链安全要求	行业标准	中国通信标准化协会	1 年
15	CD	人工智能平台供应链安	行业标准	中国通信标准化协会	1 年

		全评估方法			
16	DA、DB、DC	电信和互联网人工智能数据安全要求	行业标准	中国通信标准化协会	1年
17	DA、DB、DC	电信和互联网人工智能数据安全评估方法	行业标准	中国通信标准化协会	1年
18	DA	人工智能 安全治理 生成式人工智能服务用户数据安全能力要求	行业标准	工业和信息化部人工智能标准化技术委员会	1年
19	DB	人工智能训练数据集安全互操作要求	行业标准	工业和信息化部人工智能标准化技术委员会	2年
20	DB	生成式人工智能公共训练数据资源管理规范	行业标准	中国通信标准化协会	3年
21	DC	人工智能应用数据分类分级安全管理规范	行业标准	工业和信息化部人工智能标准化技术委员会	2年
22	DC	人工智能应用数据跨境安全防护要求	行业标准	工业和信息化部人工智能标准化技术委员会	2年
23	EA、EB	生成式人工智能算法模型安全规范	行业标准	中国通信标准化协会	2年
24	EA	电信和互联网人工智能算法安全评估指南	行业标准	中国通信标准化协会	1年
25	EA	电信和互联网人工智能算法安全要求	行业标准	中国通信标准化协会	1年
26	EB	人工智能模型开发框架安全基本要求	行业标准	工业和信息化部人工智能标准化技术委员会	1年
27	EB	多模态大模型安全基准能力测试方法	行业标准	工业和信息化部人工智能标准化技术委员会	2年
28	EB	人工智能 安全治理 安全大模型测试评价方法	行业标准	工业和信息化部人工智能标准化技术委员会	2年
29	EB	工业大模型安全评测方法	行业标准	工业和信息化部人工智能标准化技术委员会	2年
30	EB	混合模型安全要求	行业标准	工业和信息化部人工智能标准化技术委员会	2年
31	EB	多模型协同调度安全要	行业标准	工业和信息化部人工	3年

		求		智能标准化技术委员会	
32	FA	工业领域人工智能产品试验证安全要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
33	FB	智能网联汽车环境智能感知算法安全评测规范	行业标准	中国通信标准化协会	2 年
34	FB	人工智能 安全治理 行业大模型安全能力评估 第 1 部分：金融	行业标准	中国通信标准化协会	2 年
35	FB	行业大模型安全能力评估 第 2 部分：医疗	行业标准	中国通信标准化协会	2 年
36	FB	行业大模型安全能力评估 第 3 部分：政务	行业标准	中国通信标准化协会	2 年
37	FD	人工智能 安全治理 生成式人工智能视频检测服务系统能力要求	行业标准	中国通信标准化协会	2 年
38	FD	人工智能 安全治理 生成式人工智能音频检测服务系统能力要求	行业标准	中国通信标准化协会	2 年
39	FD	人工智能 安全治理 人脸识别系统防伪能力智能化分级	行业标准	中国通信标准化协会	2 年
40	FD	人工智能 安全治理 人脸识别系统注入攻击防御能力要求及评估方法	行业标准	中国通信标准化协会	2 年
41	FD	对抗环境下人脸识别系统鲁棒性安全技术要求及检测方法	行业标准	中国通信标准化协会	2 年
42	FD	人工智能模型即服务安全要求	国家标准	全国网络安全标准化技术委员会	2 年
43	FD	人工智能应用安全运营能力评估规范	国家标准	全国网络安全标准化技术委员会	2 年
44	FD	大模型检索增强知识库安全管理指南	行业标准	工业和信息化部人工智能标准化技术委员会	1 年
45	FD	互联网新技术新业务安全评估指南	行业标准	中国通信标准化协会	1 年
46	FD	互联网新技术新业务安全评估要求 基于人工智能技术的业务	行业标准	中国通信标准化协会	2 年
47	FF	人工智能 安全治理 智能体内生安全技术要求	行业标准	工业和信息化部人工智能标准化技术委员	2 年

				会	
48	FF	人工智能 安全治理 智能体服务安全技术要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
49	FF	智能体应用安全保障要求	行业标准	工业和信息化部人工智能标准化技术委员会	3 年
50	FF	智能体安全基本要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
51	FF	智能体数据接口安全要求	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
52	FF	智能体自主操作安全要求	行业标准	工业和信息化部人工智能标准化技术委员会	3 年
53	FG	互联网深度合成信息服务标识通用安全要求	行业标准	中国通信标准化协会	1 年
54	FG	反诈场景下生成式人工智能检测服务技术要求	行业标准	中国通信标准化协会	1 年
55	FG	生成式人工智能检测技术能力评测方法	行业标准	中国通信标准化协会	1 年
56	FG	生成式人工智能内容有害性分类评估规范	行业标准	中国通信标准化协会	1 年
57	FG	生成式人工智能检测系统技术要求及接口规范	行业标准	中国通信标准化协会	2 年
58	GA	生成式人工智能网络安全产品应用 成熟度要求	行业标准	中国通信标准化协会	2 年
59	GA	生成式人工智能网络安全产品应用 技术规范和评估方法	行业标准	中国通信标准化协会	2 年
60	GA	网络安全大模型技术要求	行业标准	中国通信标准化协会	1 年
61	GA	网络安全大模型测评数据集要求	行业标准	中国通信标准化协会	1 年
62	GA	网络空间安全仿真 网络安全大模型仿真平台总体要求	行业标准	中国通信标准化协会	1 年
63	GA	人工智能 安全治理 人工智能赋能恶意流量检测技术能力要求	行业标准	工业和信息化部人工智能标准化技术委员会	1 年
64	GA	人工智能赋能攻防靶场	行业标准	工业和信息化部人工	1 年

		技术能力要求		智能标准化技术委员会	
65	GA	人工智能赋能漏洞挖掘验证技术能力要求	行业标准	工业和信息化部人工智能标准化技术委员会	1 年
66	GA	网络安全大模型应用系统评测指标和方法	行业标准	中国通信标准化协会	2 年
67	GA	基于移动智能终端大模型的电信网络诈骗防范指南	行业标准	工业和信息化部人工智能标准化技术委员会	2 年
68	GB	人工智能赋能数据分类分级标准	行业标准	工业和信息化部人工智能标准化技术委员会	1 年
69	GC	人工智能 安全治理 大模型赋能内容安全能力要求与评估方法	行业标准	工业和信息化部人工智能标准化技术委员会	1 年
70	GD	人工智能 安全治理 大模型赋能业务安全能力要求与评估方法	行业标准	工业和信息化部人工智能标准化技术委员会	1 年