《安全应急大模型标准》团体标准（征求意见稿）
编制说明

一、工作简况

（一）任务来源

随着数字化转型的加速，安全应急领域面临着日益复杂的安全挑战。传统的安全防护手段已难以满足当前复杂多变的安全需求，而大模型技术的出现为提升安全应急能力带来了新的机遇。然而，当前通用大模型在安全应急领域的应用存在“通而不专”的局限性，难以满足专业场景的需求。为解决这一问题，推动安全应急大模型的专业化、规范化发展，制定《安全应急大模型标准》显得尤为必要。

本标准的立项工作是由北京广监云科技有限公司、中国信息通信研究院联合北京市科学技术研究院、船舶信息研究中心（中国船舶集团有限公司第七一四研究所）、北京中应安赫科技有限公司、北京华夏安科信息技术有限公司、联通数字科技有限公司、大家保险集团有限责任公司、防灾科技学院、华易数安科技（吉林省）有限公司、奇安信科技集团股份有限公司、亚信安全科技股份有限公司、南京赛宁信息技术有限公司、北京国科云计算技术有限公司、常州工业职业技术学院等多家单位共同发起的。牵头企业基于其在安全应急领域多年的技术积累和行业经验，结合当前行业对大模型技术的需求，于2025年4月21日向中国互联网协会团体

标准提交了《安全应急大模型标准》项目建议书。该建议书详细阐述了标准制定的必要性、目标、适用范围以及初步的技术框架。

中国互联网协会团体标准对项目建议书进行了认真评估，并于2025年4月30日正式批准立项，下达了标准编制任务。随后，牵头企业联合其他参与企业，组成了由行业专家、技术骨干等组成的起草工作组，共同开展标准的编制工作。起草工作组结合行业实际需求，参考了国内外相关标准和技术规范，经过多次研讨和论证，形成了本征求意见稿。

本标准旨在规范安全应急领域大模型的研发、部署及应用，提升模型在安全应急场景中的精准性与适配性，有效防范数据泄露、模型滥用等安全风险，促进产业协同与融合发展，为公共安全与应急管理的智能化升级提供有力支撑。

（二）制定背景

随着数字化转型的加速，安全应急领域面临着日益复杂的安全挑战。传统的安全防护手段已难以应对新型安全威胁，而大模型技术的出现为提升安全应急能力带来了新的机遇。然而，当前通用大模型在安全应急领域的应用存在“通而不专”的局限性，难以满足专业场景的需求。具体问题如下：

技术适配性不足：通用大模型虽然具备广泛的应用能力，但在安全应急领域缺乏针对性的优化，导致其在专业场景中的精准性和适配性不足。例如，在网络安全、数据保护、应

急处置等关键领域，通用大模型的性能和效果无法满足实际需求。

数据安全与隐私保护问题：大模型在数据采集、处理和存储过程中面临数据泄露和滥用的风险。尤其是在安全应急领域，数据的敏感性和重要性极高，必须通过明确的技术要求和评估方法来保障数据安全和隐私。

缺乏统一的评估标准：目前，安全应急大模型的技术水平参差不齐，缺乏统一的评估和评测体系。这不仅影响了模型的可信度，也阻碍了其在实际业务中的广泛应用。

产业协同不足：安全应急大模型的落地应用需要多方协同，包括技术研发、数据管理、应用场景开发等。然而，目前行业内缺乏有效的协同机制，导致模型与实际业务场景的融合不够紧密，难以形成新质生产力。

为解决上述问题，推动安全应急大模型的专业化、规范化发展，制定《安全应急大模型标准》显得尤为必要。本标准旨在通过明确技术框架、强化核心能力、规范应用场景及建立严格的评估体系，提升模型在安全应急场景中的精准性与适配性，有效防范数据泄露、模型滥用等安全风险，促进产业协同与融合发展，为公共安全与应急管理的智能化升级提供有力支撑。

（三）起草过程

立项与筹备阶段（[2025 年 4 月 21 日至 4 月 30 日]）

[2025 年 4 月 21 日]：由北京广监云科技有限公司、中国信息通信研究院联合北京市科学技术研究院、船舶信息研究中心（中国船舶集团有限公司第七一四研究所）、北京中应安赫科技有限公司、北京华夏安科信息技术有限公司、联通数字科技有限公司、大家保险集团有限责任公司、防灾科技学院、华易数安科技（吉林省）有限公司、奇安信科技集团股份有限公司、亚信安全科技股份有限公司、南京赛宁信息技术有限公司、北京国科云计算技术有限公司、常州工业职业技术学院等多家单位共同发起，向中国互联网协会团体标准提交了《安全应急大模型标准》项目建议书。

[2025 年 4 月 28 日]：中国互联网协会团体标准组织专家对项目建议书进行评估和论证，确认了标准制定的必要性和可行性。

[2025 年 4 月 30 日]：正式下达标准立项通知，成立标准起草工作组，明确了各参与单位和专家的职责分工。

调研与资料收集阶段（[2025 年 5 月 1 日至 6 月 26 日]）

[2025 年 5 月 1 日至 5 月 26 日]：起草工作组开展广泛的行业调研，收集国内外相关标准、技术文献、行业报告以及企业实际应用案例。

[2025 年 5 月 27 日至 6 月 26 日]：组织专家对调研结果进行分析，梳理出安全应急大模型在技术研发、部署应用、数据安全等方面的关键问题和需求。

草案编制阶段（[2025 年 6 月 27 日至 9 月 10 日]）

[2025 年 6 月 27 日至 9 月 10 日]：起草工作组根据调研结果和行业需求，初步拟定标准草案框架，明确标准的主要章节和技术内容。

[2025 年 7 月 16 日至 9 月 10 日]：组织内部讨论会，对草案框架进行详细讨论和修改，形成标准草案初稿。

征求意见阶段（[2025 年 9 月 15 日]）

[2025 年 9 月 15 日]：将标准草案初稿发送给行业内相关企业、专家和用户，广泛征求意见。

（四）起草单位、主要起草人及其所做的工作

牵头单位：中国信息通信研究院，北京广监云科技有限公司

牵头组织标准立项、起草工作，负责标准的整体规划和技术框架设计，协调各参与单位的工作。

参与单位：

[北京市科学技术研究院、大家保险集团有限责任公司]：负责安全应急大模型技术框架研究与标准条款撰写。

[船舶信息研究中心（中国船舶集团有限公司第七一四研究所）、北京中应安赫科技有限公司]：负责应用场景分析与模型评估体系的构建。

[北京华夏安科信息技术有限公司、联通数字科技有限公司、华易数安科技（吉林省）有限公司、奇安信科技集团

股份有限公司、亚信安全科技股份有限公司、南京赛宁信息技术有限公司、北京国科云计算技术有限公司、大家保险集团有限责任公司]：提供行业应用案例支持，协助验证标准的可操作性。

[防灾科技学院]：提供学术支持，参与标准的技术论证和前瞻性研究。

主要起草人及其工作

[侯卓林]

职务/职称：[总经理]

主要工作：负责标准的整体架构设计、技术框架撰写，组织协调起草工作组的工作，确保标准制定的科学性和系统性。

[蒋阿芳]

职务/职称：[总经理]

主要工作：负责标准的整体架构设计、技术框架撰写，组织协调起草工作组的工作，确保标准制定的科学性和系统性。

[王亚飞]

职务/职称：[研究员]

主要工作：安全应急大模型技术框架研究与标准条款撰写，参与标准的征求意见和反馈意见处理。

[马英轩]

职务/职称：[研究员]

主要工作：安全应急大模型技术框架研究与标准条款撰写，参与标准的征求意见和反馈意见处理。

[陈丽娜]

职务/职称：[研究员]

主要工作：安全应急大模型技术框架研究与标准条款撰写，参与标准的征求意见和反馈意见处理。

[谷春野]

职务/职称：[技术经理]

主要工作：安全应急大模型技术框架研究与标准条款撰写，参与标准的征求意见和反馈意见处理。

[秦绪坤]

职务/职称：[高级工程师]

主要工作：负责应用场景分析，构建模型评估体系，撰写相关标准条款，参与标准的征求意见和反馈意见处理。

[张英香]

职务/职称：[研究员]

主要工作：提供行业应用案例支持，协助验证标准的可操作性，参与标准的征求意见和反馈意见处理。

[高竞秀]

职务/职称：[解决方案工程师]

主要工作：负责应用场景分析，撰写相关标准条款，参

与标准的征求意见和反馈意见处理。

[丛磊]

职务/职称：[技术经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[王冉]

职务/职称：[技术经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[刘岩]

职务/职称：[技术经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[杨婷]

职务/职称：[技术经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[颜科]

职务/职称：[技术经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[邹立刚]

职务/职称：[技术经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[崔鹏飞]

职务/职称：[经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[罗华伟]

职务/职称：[经理]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[郝亚平]

职务/职称：[研究院]

主要工作：提供行业应用案例支持，协助验证标准的可操作性。

[吴燕雄]

职务/职称：[教授]

主要工作：提供学术支持，参与标准的技术论证和前瞻性研究，确保标准的技术先进性和科学性。

二、编制原则、主要内容及其确定的来源和依据

（一）编制原则

1. 科学性与先进性

标准的制定以科学为基础，充分考虑当前安全应急大模型技术的最新发展水平，确保标准内容具有前瞻性和技术先进性。引用国内外最新的研究成果和技术规范，确保标准能够适应未来技术的发展趋势。

在技术框架和评估方法的设计中，采用科学的理论和方法，确保标准的技术条款具有可验证性和可操作性。

2. 实用性与可操作性

标准紧密结合行业实际需求，确保其内容具有实用性和可操作性。标准条款应简洁明了，易于理解和执行，避免过于复杂或模糊的表述。

通过广泛的行业调研和实际案例分析，确保标准能够解决实际问题，满足企业在安全应急大模型研发、部署和应用中的具体需求。

3. 安全性与可靠性

标准特别强调数据安全和隐私保护，明确模型在数据采集、处理、存储等环节的安全规范，防范数据泄露和滥用风险。

确保模型在基础网络安全、数据安全、内容安全和业务安全四个核心领域的功能性、有效性和合规性，保障系统的可靠性和稳定性。

4. 规范性与一致性

标准的制定遵循国家相关法律法规和政策要求，与现有

的国家标准、行业标准保持一致，避免冲突或重复。

在标准的格式、术语定义、符号使用等方面，遵循标准化的基本要求，确保标准的规范性和统一性。

5. 开放性与兼容性

标准在制定过程中充分考虑了与其他相关标准和技术的兼容性，确保安全应急大模型能够与其他系统和平台无缝对接。

标准内容具有一定的开放性，鼓励技术创新和行业合作，为未来的技术发展和标准修订留出空间。

6. 产业协同与用户需求导向

标准的制定以推动产业协同和满足用户需求为导向，促进安全应急大模型与实际业务场景的深度融合，提升公共安全与应急管理的智能化水平。

广泛征求行业内外的意见和建议，确保标准能够反映各方利益，推动行业的健康发展。

（三）主要内容及其确定依据

《安全应急大模型标准》的主要内容包括技术框架、安全能力、应用场景和评估与评测体系。技术框架明确了安全应急大模型的总体架构，基于通用大模型进行安全领域的专项训练与优化，构建涵盖通用安全能力与原子能力的技术体系。安全能力条款确保模型在基础网络安全、数据安全、内容安全和业务安全四个核心领域的功能性、有效性和合规性，

保障用户的数据隐私、系统稳定性以及安全防护能力。应用场景覆盖监测预警、风险评估、事件研判、态势分析等典型场景，如告警研判、漏洞挖掘、应急处置辅助等，推动模型在实际业务中的落地应用。评估与评测体系则建立科学的评估方法与指标，规范模型的研发测试、部署应用及第三方评测流程，保障模型的有效性与可信度。

这些内容的确定依据是多方面的。技术框架的设计基于当前安全应急领域对大模型技术的实际需求，参考了国内外相关技术文献、行业报告以及实际应用案例，通过广泛的行业调研和专家论证确定。安全能力的条款参考了国家相关法律法规（如《网络安全法》《数据安全法》等）以及国际标准（如 ISO/IEC 27001 等），结合安全应急领域的实际需求制定。应用场景的覆盖范围是通过对行业内多家企业的调研以及对实际应用案例的分析确定的，确保涵盖安全应急领域的关键环节。评估与评测体系的建立参考了国内外现有的模型评估标准和技术规范，结合安全应急领域的特殊需求制定，为政府、企业等用户提供可信赖的技术支撑。

（四）修订前后技术内容的对比（修订项目时应有这个内容）

无。

三、标准验证情况

1. 验证目的

标准验证的目的是确保《安全应急大模型标准》的条款具有科学性、可操作性和有效性，能够满足安全应急领域大模型研发、部署和应用的实际需求。通过验证，进一步完善标准内容，提高标准的实用性和可信度。

2. 验证方法

内部测试：起草工作组在牵头单位和参与单位的支持下，对标准草案中的技术框架、安全能力、应用场景和评估体系进行了内部测试。测试内容包括模型的训练、部署、评估等环节，确保标准条款的可操作性。

企业试点：选择行业内具有代表性的企业进行标准试点应用。试点企业根据标准要求，对现有的安全应急大模型进行优化和调整，并反馈实际应用中的问题和建议。

专家评审：组织行业专家对标准草案进行评审，专家们从技术、安全、应用等多个角度对标准内容进行评估，提出修改意见和建议。

用户反馈：广泛征求用户意见，通过问卷调查、座谈会等方式，收集用户对标准草案的反馈，确保标准能够满足实际需求。

3. 验证结果

内部测试结果：内部测试表明，标准草案中的技术框架和评估体系具有较高的科学性和可操作性。模型在基础网络安全、数据安全、内容安全和业务安全四个核心领域的表现

符合预期，能够有效防范数据泄露和模型滥用等安全风险。

企业试点结果：试点企业反馈，标准的应用能够显著提升安全应急大模型的性能和可靠性。通过标准的指导，企业在模型优化、数据管理、安全防护等方面取得了显著进展，模型在实际业务场景中的应用效果得到了用户的认可。

专家评审意见：专家评审一致认为，标准内容全面、科学，具有较强的指导性和实用性。专家们提出了一些完善建议，起草工作组已根据这些建议对标准草案进行了修改和完善。

用户反馈结果：用户反馈表明，标准能够满足行业需求，特别是在数据安全和隐私保护方面，用户对标准的条款给予了高度评价。用户建议进一步细化应用场景的描述，以便更好地指导实际工作。

4. 验证结论

通过内部测试、企业试点、专家评审和用户反馈等多方面的验证，标准验证情况表明，《安全应急大模型标准》具有较高的科学性、可操作性和有效性。标准的实施能够有效提升安全应急大模型的性能和可靠性，满足行业实际需求。起草工作组将根据验证结果进一步完善标准内容，确保标准的高质量和实用性。

四、与国际、国外同类标准技术内容的对比情况，或者与测试的国外样品、样机的有关数据对比情况

国际方面，大模型安全标准研究已取得阶段性成果。2024年4月，世界数字技术院（WDTA）在联合国科技大会期间发布《生成式人工智能应用安全测试标准》和《大语言模型安全测试方法》两项国际标准，由OpenAI、蚂蚁集团、谷歌、微软等数十家机构联合编制。前者为生成式AI应用的安全性评估提供全生命周期框架，后者则针对大语言模型提出安全风险分类、攻击分级及测试方法，推动了全球大模型安全测试的规范化发展。

该标准项目将充分借鉴国际先进经验，参考WDTA发布的大模型安全测试标准，在风险分类、攻击检测及测试方法等方面进行技术衔接。同时，结合国内安全应急领域的特殊需求，如矿山安全、城市内涝预警等场景化应用，对国际标准进行适应性调整。例如，在数据安全要求上，将严格遵循国内数据保护法规，强化敏感数据分级保护；在应急响应流程中，融入中国应急管理体系的标准化流程，确保标准既具国际视野又符合本土实践。

五、采用国际标准的情况

无。

六、与有关的法律、法规和相关标准的关系

《安全应急大模型标准》的制定严格遵循国家相关法律法规，确保与现有法律框架保持一致。本标准参考了《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中

华人民共和国个人信息保护法》等法律法规，明确了数据安全、隐私保护、网络安全等方面的要求，确保标准的实施不会与现行法律相冲突。

同时，本标准与现有国家标准和行业标准保持高度一致。例如，参考了《信息安全技术 网络产品安全漏洞管理规范》（GB/T 39204-2020）、《信息安全技术 个人信息安全规范》（GB/T 35273-2020）等国家标准，以及《安全大模型能力要求与评估方法》等系列规范。在技术框架、安全能力、应用场景和评估体系等方面，本标准与上述标准相互补充，共同构建了安全应急领域大模型的标准化体系。

此外，该标准与国内现有应急管理及人工智能领域标准形成互补与协同。一方面，与应急管理部《“十四五”应急管理标准化发展计划》中关于安全生产、消防救援、减灾救灾等标准体系相衔接，尤其在监测预警、数据治理、应急指挥等环节强化智能化支撑。另一方面，参考中国信通院《安全大模型能力要求与评估方法》，在技术架构、安全能力维度等方面保持一致性，避免重复建设。同时，标准将推动安全应急大模型与地方标准（如北京市在隐患识别、应急知识服务等场景的探索）的协同发展，形成从国家到地方的多层次标准体系。

在制定过程中，本标准注重与现有标准的协同性，避免重复和冲突。对于已有明确规定的领域，本标准不再另行制

定，而是直接引用相关标准的条款。对于尚未涵盖的内容，本标准进行了补充和完善，确保标准体系的完整性。通过这种方式，本标准旨在为安全应急大模型的研发、部署和应用提供全面、系统的指导，推动行业的规范化发展。

七、重大分歧意见的处理经过和依据

无。

八、涉及专利的有关说明

目前公开资料未明确提及该标准项目存在知识产权争议。国际标准编制过程中，由多家企业联合参与，通过协作机制处理知识产权问题；国内标准制定亦注重产学研结合，例如中国信通院联合百度、科大讯飞等企业共同起草规范，通过协商明确知识产权归属。在《安全应急大模型》标准制定过程中，需延续这一模式，确保参与单位的知识产权得到充分尊重，同时遵循开源技术及引用标准的合规性要求，避免潜在法律风险。建议在标准制定过程中建立知识产权评估机制，确保技术内容的合法性与独立性。

九、其他应当说明的事项

无。