《数字安全大模型应用能力成熟度评估规范》团体标准（征求意见稿）

编制说明

一、工作简况

（一）任务来源

随着数字化转型的深入推进，人工智能技术特别是大模型在全球范围内快速发展，并逐步在数字安全领域形成广泛应用共识，通过结合安全知识增强训练与行业场景适配，数字安全大模型已在安全运营、威胁检测、漏洞分析及事件响应等多个关键环节中展现出重要价值；在国际层面，相关安全大模型产品已在多类安全场景中开展实践探索，在国内，已有30家以上安全厂商推出数字安全大模型产品并在政企及关键信息基础设施等生产环境中落地应用，但与此同时仍存在能力评价体系不统一、安全防护能力缺乏量化标准、应用成熟度边界不清晰及业务融合程度不足等问题，因此为规范数字安全大模型应用能力评估体系建设，推动行业从分散探索向体系化、标准化与规模化应用演进，制定《数字安全大模型应用能力成熟度评估规范》显得尤为必要。

本标准的立项工作是由中国信息通信研究院、南方电网互联网服务有限公司、国网冀北电力有限公司智能配电网中心、大连银行股份有限公司、北京安普诺信息技术有限公司、闪捷信息科技有限公司、亚信安全科技股份有限公司、云南

易数科技有限责任公司等多家单位共同发起的。牵头企业基于其在各自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[2024 年 7 月]向中国互联网协会团体标准提交了《数字安全大模型应用能力成熟度评估规范》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[2024 年 7 月]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。

（二）制定背景

随着数字化转型的持续深入推进，数字安全领域面临的安全威胁呈现出复杂化、智能化与多样化发展趋势，传统安全防护与分析手段在应对高级持续性攻击、复杂攻击链识别及大规模安全事件处置等方面逐渐暴露出能力瓶颈。与此同时，大模型技术的快速发展为提升数字安全整体能力提供了新的技术路径，使安全运营、威胁检测与智能研判等任务具备了更高程度的自动化与智能化基础。

然而，当前通用大模型在数字安全领域的应用仍存在明显的“通用能力强、专业能力弱”的结构性问题，难以满足

安全运营、漏洞分析、攻击溯源及事件响应等专业化场景需求。具体来看，一是技术适配性不足，通用大模型虽具备较强语言理解与生成能力，但在安全专业知识建模与复杂安全任务处理方面缺乏针对性优化；二是数据安全与隐私保护风险突出，大模型在数据采集、训练及推理过程中涉及大量敏感安全数据，缺乏统一规范时易产生数据泄露与滥用风险；三是缺乏统一的能力评估标准，不同厂商与产品之间在能力定义、评估方法与成熟度划分上存在差异，影响行业的横向对标与可信评价。

同时，数字安全大模型的落地应用高度依赖技术研发、数据治理与安全运营等多方协同，但当前行业整体协同机制仍不完善，模型能力与实际业务场景之间融合深度不足，制约了规模化应用与价值释放。在此背景下，为系统规范数字安全大模型应用能力评估方法，构建统一的成熟度评价体系与分级标准，推动行业从分散探索向体系化建设与规范化发展演进，制定《数字安全大模型应用能力成熟度评估规范》显得尤为必要。

（三）起草过程

立项与筹备阶段（[2024 年 5 月至 2024 年 7 月]）

[2024 年 7 月]：由中国信息通信研究院、南方电网互联网服务有限公司、国网冀北电力有限公司智能配电网中心、大连银行股份有限公司、北京安普诺信息技术有限公司、闪

捷信息科技有限公司、亚信安全科技股份有限公司、云南易数科技有限责任公司多家单位共同发起，向中国互联网协会团体标准提交了《数字安全大模型应用能力成熟度评估规范》项目建议书。

[2024 年 8 月]：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。

[2024 年 9 月]：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（[2024 年 10 月至 2024 年 12 月]）

[2025 年 1 月]：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 2 月]：组织专家对调研结果进行分析，梳理出数字安全大模型应用能力成熟度评估规范在技术研发、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 3 月至 2025 年 10 月]）

[2025 年 11 月]：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 12 月]：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。

[2025 年 12 月]至今：目前正处于征求意见阶段

（四）起草单位、主要起草人及其所做的工作

牵头单位：由中国信息通信研究院牵头组织标准立项、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：

[北京安普诺信息技术有限公司、亚信安全科技股份有限公司]：负责数字安全大模型应用能力成熟度评估规范技术框架研究与标准条款撰写。

[闪捷信息科技有限公司、云南易数科技有限责任公司]：负责应用场景分析与模型评估体系的构建。

[南方电网互联网服务有限公司、国网冀北电力有限公司智能配电网中心、大连银行股份有限公司、北京奕华科技有限公司]：提供行业应用案例支持，协助验证标准的可操作性。

主要起草人及其工作

[马英轩]

职务/职称：[研究员]

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。

[侯洪磊]

职务/职称：[研究员]

主要工作：负责协助标准架构设计、组织协调起草工作组的工作，收集整合行业应用案例，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：负责数字安全大模型应用能力成熟度评估规范技术框架研究与标准条款撰写，确保标准制定的科学性和系统性。

[张涛]

职务/职称：[CEO]

主要工作：数字安全大模型应用能力成熟度评估规范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。

[杨婷]

职务/职称：[研究员]

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。

[苏伟华]

职务/职称：[研究员]

主要工作：负责应用场景分析与模型评估体系的构建。

[赵吕]

职务/职称：[技术总负责人]

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。

[苏丕云]

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。

[桂媛]

职务/职称：[高级工程师]

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。

[夏武]

职务/职称：[高级工程师]

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。

[张润学]

职务/职称：[高级政工师

主要工作：负责应用场景分析与模型评估体系的构建。

[陈志敏]

职务/职称：[工程师]

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。

[张子超]

职务/职称：[信息科副总经理]

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。

[胡珏]

职务/职称：[项目经理]

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。

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，充分考虑当前数字安全大模型应用能力成熟度评估规范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，满足企业在数字安全大模型应用能力成熟度评估规范研发、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，确保数字安全大模型应用能力成熟度评估规范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，促进数字安全大模型应用能力成熟度评估规范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

《数字安全大模型应用能力成熟度评估规范》的主要内容包括技术框架、安全能力、应用场景及评估与评测体系四个方面。技术框架明确了数字安全大模型应用能力成熟度评估的总体结构，基于通用大模型能力扩展与安全领域专项优化，构建覆盖产品基础能力、模型工程能力、业务融合能力、安全治理能力、组织运营能力及持续优化能力的能力域体系，并通过 L1 至 L5 五级成熟度模型刻画能力演进路径与阶段特征。安全能力条款重点围绕数字安全大模型在全生命周期中的安全防护要求展开，涵盖输入安全防护、输出内容控制、训练数据安全管理及模型运行安全保障等内容，同时覆盖基础网络安全、数据安全、内容安全与业务安全等核心维度，用于保障模型在实际应用中的安全性、可靠性与合规性。

这些内容的确定依据是多方面的。一方面，技术框架设计基于数字安全领域对大模型应用的实际需求，参考国内外

相关技术文献、行业研究报告及典型应用案例，并通过广泛行业调研与专家论证形成统一能力体系结构；另一方面，安全能力条款结合《网络安全法》《数据安全法》《个人信息保护法》等法律法规要求，并参考 ISO/IEC 27001 等国际信息安全管理体系标准，同时结合数字安全场景的实际特点进行本地化适配设计，从而确保标准内容具备合规性与可操作性。

（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《数字安全大模型应用能力成熟度评估规范》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，对现有的数字安全大模型应用能力成熟度评估规范进行优化和调整，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数字安全大模型应用能力成熟度评估规范的性能和可靠性。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，《数字安全大模型应用能力成熟度评估规范》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数字安全大模型应用能力成熟度评估规范的性能和可靠性，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

国际层面，Google 发布了自研的网络安全大模型 Sec-PaLM2，该模型针对安全应用场景进行了微调，并整合了大量威胁情报数据；微软发布了 Security Copilot，该产品结合了 GPT-4 与微软的安全数据，具备恶意脚本分析等多种安全能力。总体来看，国际相关研究正加速推动大模型在网络安全领域的应用落地，并不断拓展安全分析与自动化响应能力边界。

国内层面，诸多安全厂商已发布自主研发的数字安全大模型产品，并在安全运营、威胁检测等场景开展应用实践。同时，中国信息通信研究院联合多家科技厂商发布《大模型安全研究报告（2024）》，提出了大模型安全框架，为行业技术发展提供重要参考。整体来看，安全大模型正朝着实战应用方向加速发展，并持续探索更多应用场景，以提升自动化安全运营的效率与质量。

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

本标准在制定过程中充分遵循国家相关法律法规要求，与现行网络安全、数据安全及人工智能治理体系保持一致，确保标准内容的合规性与适用性。在法律法规层面，本标准主要依据《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》等相关法律要求，重点围绕数据安全保护、个人信息合规使用及网络安全管理等方面提出能力约束与评估要求，确保数字安全大模型在应用过程中满足国家法律法规的基本要求。

在相关标准方面，本标准参考并衔接GB/T 37988《信息安全技术 数据安全能力成熟度模型》、GB/T 43439《数字化转型成熟度模型与评估》、GB/T 25069《信息安全技术 术语》以及ISO/IEC 42001（人工智能管理体系）、ISO/IEC 23894

（人工智能风险管理指南）等国内外标准体系，同时与中国信息通信研究院牵头制定的T/CCSA 561.1—8《面向行业的大规模预训练模型通用要求》系列标准保持协调一致，在模型能力要求、安全治理体系及风险管理方法等方面实现衔接与互补。

本标准重点面向数字安全大模型应用能力成熟度评估场景，侧重于能力分级、应用成熟度评价及安全治理能力评估，是对现有通用AI治理标准及数据安全能力标准在“数字安全垂直领域”的进一步细化与扩展，不替代已有国家标准与行业标准，而是在其基础上形成补充性、应用性与评估性规范，为数字安全大模型的建设、应用与第三方评估提供统一依据。

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，本标准旨在为数字安全大模型应用能力成熟度评估规范的研发、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北京安普诺信息技术有限公司、亚信安全科技股份有限公司等企业共同起草规范，通过协商明确知识产权归属。在《数字安全大模型应用能力成熟度评估规范》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。