

《证券期货业智能模型安全检测要求》

编制说明

《证券期货业智能模型安全检测要求》标准起草组

二〇二六年十月

一、背景及意义

1.背景

证券期货业智能模型在风险管理、综合决策和合规审查等业务环节的应用不断深化。模型上线前及运行中的安全状态需要通过规范化检测进行验证。现有通用标准缺乏面向证券期货业业务场景的检测框架、流程、方法和通过准则，可能导致检测口径不一致、结果难以复现。

2.目的和意义

本项目针对模型风险管理普遍存在的重原则、轻检测以及模型上线前缺少统一可量化检测方法的问题，以《证券期货业智能模型安全基本要求》为依据，规定了证券期货业智能模型安全检测的目标、流程、方法和通过要求。检测内容涵盖数据安全、模型有效性、安全性、鲁棒性、隐私性、公平性、可解释性、持续维护能力和安全监测与响应等维度，供金融机构、第三方检测机构及相关管理单位在智能模型开发验证、上线前检测、运行期复测等过程中参考使用，旨在将模型安全管理原则和基本要求转化为统一有效的评估手段，客观评估模型安全状况与风险，实现动态监测与持续验证，是落实模型安全基本要求、防控模型风险的关键技术支撑。

二、工作简况

1.任务来源

标准制定任务源于证券期货行业对智能模型安全统一规范的实际需求。随着人工智能技术在风控、投研、量化交易、智能客服等业务环节的深入应用，模型偏差、算法歧视、数据泄露、性能漂移等模型风险日益凸显，行业亟需统一、可执行、可验证的智能模型安全标准。研制工作依托国家重点研发计划项目“金融数据合成与智能模型风险监测关键技术及应用”（项目编号：2023YFC3304700）课题三“金融智能模型风险评估与监管技术研究”（课题编号：2023YFC3304703）开展。该课题将“形成智能模型安全领域的行业标准送审稿”列为重要预期成果，本文件即为该课题在标准制定方面的核心产出，其安全要求框架与指标体系直接来源于课题在模型有效性、安全性、鲁棒性、隐私性、公平性、可解释性及全生命周期风险监测与治理等维度的研究成果。

本项目为金融行业标准制定项目，由北京航空航天大学牵头，申万宏源证券有限公司、华金证券股份有限公司、金圆统一证券股份有限公司、格林基金管理有限公司参与起草。主要起草人如下：

表 1 主要起草人

序号	姓名	单位	主要工作内容/职责
1	吴俊杰	北京航空航天大学	编写组组长，负责项目整体管理与全文统稿，牵头编写第 1~3 章（范围、规范性引用文件、术语和定义）及第 4 章总体安全要求
2	韩潇	北京航空航天大学	负责第 5 章数据存储与取用安全要求、第 6 章模型有效性要求的编写
3	左源	北京航空航天大学	负责第 7 章模型安全性要求、第 8 章模型鲁棒性要求的编写
4	朱天宇	北京航空航天大学	负责第 9 章模型隐私性要求、第 10 章模型公平性要求的编写
5	部慧	北京航空航天大学	负责第 11 章模型可解释性要求、第 12 章模型持续维护能力要求、第 13 章模型安全监测与响应机制要求的编写，兼顾组织协调
6	王忞思	申万宏源证券有限公司	负责第 4~13 章行业适用性审阅，提供证券业务场景与风控实践案例支持
7	王叶	申万宏源证券有限公司	负责第 5~10 章行业适用性审阅，提供数据与算法场景支持
8	邓科夫	华金证券股份有限公司	负责第 4~13 章行业适用性审阅，提供业务合规实践反馈
9	张煜	金圆统一证券股份有限公司	负责第 6~13 章行业适用性审阅，提供业务场景反馈
10	祝荣标	金圆统一证券股份有限公司	负责第 4~9 章行业适用性审阅，提供业务场景反馈
11	马文杰	格林基金管理有限公司	负责第 4~13 章行业适用性审阅，提供基金业务场景与量化模型实践反馈

项目组围绕本项目设置技术、协调和评审等工作机制，分别承担技术内容编写与验证、沟通与意见征集、内部质量把关与专家对接等工作。上述组织安排以项目立项后的实际工作机制为准。

2.主要工作过程

（一）项目启动

2026年3月，标准编写项目启动。

（二）成立标准编制工作组

2026年3月，标准起草小组成立并起草形成标准草案。

（三）标准立项下达

2026年4月，参加并完成标准立项答辩。

2026年5月，工作组提交草案，全国金融标准化技术委员会证券分技术委员会进行审查。起草组根据标准立项审查会要求进行修改，形成初稿。

2026年6月，工作组提交初稿进行技术审查，根据函审意见进行修改。

2026年7月，召开标准研讨会，行业专家和标准起草单位代表进行研讨，对标准适用范围、框架细节等方面提出要求。随后，向全国金融标准化技术委员会证券分技术委员会提出《证券期货业智能模型安全要求规范》标准立项申请，再次召开标准研讨会，对初稿进行讨论。

2026年7月，根据研讨会要求进行修改，形成修改稿，向行业内各大单位征求意见。

（四）形成征求意见稿

2026年7-8月，小范围征求证券企业行业部分业界专家意见，根据反馈意见进行修改完善，形成第一稿修订稿。

2026年8月，小范围征求部门监管部门和直属机构的意见，并根据意见反馈情况进行修改完善，形成第二轮修订稿。

三、编制主要内容

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草

规则》的规定起草。在技术内容确定上遵循依法合规、风险导向、全生命周期、技术与管理并重、可实施和可验证的原则：检测要求与网络安全、数据安全领域法律法规及监管规则相衔接；以模型应用中的主要风险为检测重点，检测强度与受检对象的风险水平相匹配；检测覆盖模型从数据、研发到运行维护各环节的安全要求；既采用技术测试手段，也兼顾文档审查、流程核查等管理性核查方法；各项检测要求力求明确具体、可操作、可复现，确保检测结果客观公正。

标准内容在衔接证券期货业网络安全和金融数据安全要求、避免与既有标准重复交叉的基础上，与《基本要求》提出的各项安全要求相对应，将其转化为可检测、可验证的检测项和判定准则；同时结合智能模型在风险管理、投研分析、量化交易、合规审查、智能客服等业务环节的应用特点和风险差异确定检测重点，突出针对性和适用性。

标准草案包括范围、规范性引用文件、术语和定义、总体要求，以及数据存储与取用安全、模型有效性、安全性、鲁棒性、隐私性、公平性、可解释性、持续维护能力和安全监测与响应机制等检测要求。总体要求明确检测对象与范围、检测框架、检测过程、检测方法以及检测结果与报告：界定适用对象和检测边界，构建覆盖各安全维度的检测框架，规定从检测准备、方案制定、检测实施到结果判定的过程要求，明确技术测试与文档审查、访谈核查等方法的适用情形，以及检测结果的判定规则和报告内容要求。各检测维度章节统一按检测目的、检测流程和通过要求组织：检测目的说明该维度所要验证的安全目标，检测流程规定具体检测步骤和实施方式，通过要求明确判定准则，并与《基本要求》中的对应条款相衔接，形成“安全要求—检测验证”配套贯通的标准体系。

四、主要试验（或验证）情况分析

标准草案已形成覆盖数据、有效性、安全性、鲁棒性、隐私性、公平性、可解释性、持续维护和安全监测响应的要求框架。

在标准发布后的落地实施方面，拟按“宣贯培训—试点先行—工具配套—全面实施—跟踪评估”的路径推进：

一是开展宣贯培训。由归口单位会同行业主管部门组织面向证券期货经营机构、模型提供方和检测机构的系列宣贯培训，配套编制标准解读材料、实施指引和常见问题解答，帮助各方准确理解并统一执行口径。

二是先行试点验证。选择风险管理、量化交易、智能客服等典型业务场景中基础较好、代表性强的机构和模型先行试点，验证各项要求的可操作性与适用性，沉淀可复制、可推广的实施方案。

三是完善配套支撑。推动研制配套的自评估工具和实施指南，支持机构对照标准开展模型安全自查与差距分析；推动检测机构依据本文件及《证券期货业智能模型安全检测要求》建立检测细则与测试基准，形成“要求—检测”配套落地能力。

四是分步全面实施。建议设置 6 至 12 个月过渡期：新研发、新上线的模型及应用自标准实施之日起直接适用；存量模型结合重要程度和风险状况制定整改计划、分步达标，重大风险项优先整改。实施中推动机构将标准要求嵌入模型研发、上线审核、运行监控和事件处置等内部制度与业务流程。

五是跟踪评估、动态完善。建立实施效果跟踪评估机制，结合试点情况、案例积累和技术演进定期评估标准的适用性，适时组织修订。

通过统一行业智能模型安全要求，有助于降低模型失效、数据泄露、违规输出和安全攻击等风险，减少模型整改与重复评估成本，提升证券期货业务创新的安全性、稳健性和投资者权益保护水平。

本文件面向证券公司、期货公司、基金公司、信息技术服务机构和模型提供方。实施后可为相关机构建立智能模型安全治理、风险分级、数据保护、模型研发、部署运行和持续监测机制提供基础参考，并为配套检测评估活动提供要求依据。

经济效益方面，本文件为证券期货业机构开展智能模型安全管理与检测提供统一依据，避免各机构自行摸索标准、重复建设检测能力，降低模型研发、上线评估与合规审查成本。依据标准对市场风险监测、交易欺诈检测、异常配资识别、账户异常监测等场景中应用的智能模型实施安全检测，可及早发现模型缺陷与安全隐患，减少因模型失效、模型滥用等引发的直接经济损失和整改返工成本，按单家机构减少一次专项检测外采或重复建设测算，每年可节约数十万元级投入。提升模型研发迭代效率与风险识别能力，相关效益将随标准在行业内的采标应用逐步显现。

社会效益方面，本文件的实施有助于提升行业对模型鲁棒性不足、对抗攻击、异常市场行为等风险的识别与主动防控能力，降低系统性金融风险和投资者权益损害，为“人工智能+金融”稳健落地提供安全保障；推动智能模型应用实现全过程可测、可控、

可审计，为监管机构开展模型风险监管和穿透式监管提供统一的技术依据；完善证券期货业智能模型安全标准体系，助力行业形成自主可控的智能模型安全治理能力。

总体而言，本文件将在合规可控的前提下，促进智能模型的规范应用与有序创新，提升行业整体模型安全水平，为金融强国建设和数字经济高质量发展提供基础支撑。

五、与有关的现行法律、法规和国家标准的关系

本文件与现行法律、法规和监管要求保持衔接，不替代相关合规义务。

本文件的研制对相关法律法规的落地实施具有正向促进作用。一方面，本文件将《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》《中华人民共和国证券法》《中华人民共和国期货和衍生品法》等法律以及金融监管部门有关网络安全、数据安全、投资者保护等监管规则中的原则性要求，结合证券期货业智能模型应用特点细化为可执行、可验证的行业标准要求，为行业机构落实法定合规义务提供了具体路径和技术抓手，有利于降低合规执行的不确定性。另一方面，本文件填补了证券期货业智能模型安全领域行业标准的空白，为监管部门开展模型风险监管和穿透式监管提供统一的技术依据，有助于提升监管效能、减少重复监管成本，促进"人工智能+金融"在法治轨道上稳健发展。

标准草案引用 GB/T 41479—2022《信息安全技术 网络数据处理安全要求》、GB/T 42888—2023《信息安全技术 机器学习算法安全评估规范》、JR/T 0060—2021《证券期货业网络安全等级保护基本要求》、JR/T 0223—2021《金融数据安全 数据生命周期安全规范》、GB/T 37964—2019《信息安全技术 个人信息去标识化指南》等文件。本文件在上述通用数据安全、算法安全、网络安全和金融数据安全要求的基础上，结合证券期货业智能模型应用特点形成专项规定。

六、重大分歧意见的处理经过和依据

在本文件的编制过程中，工作组成员对标准的内容进行了充分的研究和探讨，在编制过程中没有出现重大分歧意见。

七、贯彻标准的要求和建议措施

（一）组织措施

1.系统开展宣贯培训。建议在标准发布后，由标准归口单位会同行业主管部门组织系列宣贯培训，面向证券期货经营机构、模型提供方和检测机构等各方主体，重点解读标准的适用范围、核心术语定义、重点条款内涵及具体实施方法。宣贯工作可采取专题宣讲、线上课程、实务研讨、案例分析等多种形式，并配套编制标准解读材料、实施指引和常见问题解答，帮助各方准确理解标准要求，减少理解偏差和执行走样。

2.建立健全实施协调机制。推动证券期货经营机构明确标准落地的牵头部门和职责分工，建立科技、合规、风控、业务等多部门协同工作机制；加强经营机构与模型提供方、检测机构之间的沟通衔接，明确各方在模型研发、交付、检测、运维等环节的责任边界，形成全链条贯通的落实合力。

3.将标准要求嵌入内部管理制度与业务流程。推动相关机构对照标准开展差距分析，将适用要求纳入模型研发、上线审核、运行监控和事件处置等全生命周期各环节：在研发阶段落实数据、算法和安全要求；在上线环节建立审核把关机制；在运行阶段完善持续监控和记录留存；在事件处置环节健全响应、报告与整改流程，实现标准要求与内部制度的有机衔接。

4.加强监督评估与队伍建设。建议行业主管部门、行业协会将标准落实情况纳入检查评估范围，督促机构整改落实；鼓励机构培养既懂证券期货业务又懂模型技术的复合型人才，为标准实施提供组织和人才保障。

（二）技术措施

1.构建覆盖模型全生命周期的技术管控能力。推动机构在模型研发阶段落实数据质量与合规管控、算法安全评估等要求；在上线前开展功能、性能、安全性等测试验证；在运行阶段建立日志记录、内容监测、异常告警等监控机制；针对可能出现的模型失效、生成违法违规内容等情形，制定技术层面的应急处置预案（如熔断、降级、回退措施），确保风险可控。

2.完善检测验证方法与工具。支持检测机构依据标准细化检测细则和评价方法，建设标准化的测试基准、测试数据集和检测工具链，提升检测结果的科学性、一致性和可复现性；鼓励第三方检测与机构自测相结合，形成多元互补的质量验证体系。

3.强化文档管理与证据留存。推动机构按标准要求完整记录模型研发过程、训练数据来源、测试结果、运行日志和事件处置情况，做到过程可追溯、结果可核查，为监管检查和责任认定提供技术支撑。

4.跟踪技术演进、持续迭代工具方法。针对大模型技术快速迭代的特点，鼓励产检研各方持续研究新的评测方法和风险防控技术，及时更新测试基准与管控手段，保持标准实施与技术发展相适应。

（三）过渡办法

1.设置合理过渡期。考虑到机构现有系统和流程的改造需要一定周期，建议标准发布后设置合理的过渡期（如6个月至12个月），过渡期内允许存量系统对照标准开展自查整改；过渡期结束后，新上线系统和新增应用应全面执行标准要求。

2.实行存量与增量分类管理。对新研发、新采购、新上线的模型及应用，自标准实施之日起直接适用；对已在运行的存量应用，允许机构结合系统重要程度和风险状况，制定整改计划、分步达到标准要求，重大风险项应优先整改。

3.坚持试点先行、以点带面。选择部分基础较好、代表性强的机构和典型应用场景开展试点，结合试点应用和案例积累，总结形成可复制、可推广的实施经验，再逐步在全行业推开，降低整体实施成本和风险。

4.动态评估、适时修订。建立标准实施效果跟踪评估机制，结合试点应用情况、典型案例积累和大模型技术发展，定期评估标准的适用性和有效性；对实施中发现的问题和技术演进带来的新需求，及时组织研究，适时启动标准修订，保持标准的先进性和可操作性。

八、标准属性的建议

鉴于本文件的内容未涉及强制性标准或强制性条文的内容及要求，建议本文件作为推荐性金融行业标准发布。

九、废止现行相关标准的建议

本文件为首次制定，不存在废止现行相关标准的问题。

十、其他说明事项

无。